HETEROLOGOUS PRODUCTION OF BIOACTIVE TERPENES AND DISCOVERY OF NOVEL NATURAL PRODUCTS FROM MEMBERS OF THE GENUS *KITASATOSPORA*

A Thesis

Submitted to the Graduate Faculty

in Partial Fulfilment of the Requirements

for the Degree of

Doctor of Philosophy

in the Department of Biomedical Sciences

Faculty of Veterinary Medicine

University of Prince Edward Island

Jennifer Cuillerier Arens

Charlottetown, P. E. I.

May, 2015

THESIS/DISSERTATION NON-EXCLUSIVE LICENSE

Family Name: Arens	Given Name, Middle Name (if applicable): Jennifer Carolyne (Cuillerier)
Full Name of University: University of Prince Edward Island	
Faculty, Department, School: Biomedical Sciences, Faculty of Veterinary Medicine	
Degree for which thesis/dissertation was presented: Doctor of Philosophy	Date Degree Awarded: May 25th 2015
Thesis/dissertation Title: Heterologous production of bioactive terpenes and discovery of novel natural products from members of the genus *Kitasatospora*	
Date of Birth. It is **optional** to supply your date of birth. If you choose to do so please note that the information will be included in the bibliographic record for your thesis/dissertation. April 26th 1987	

In consideration of my University making my thesis/dissertation available to interested persons, I,

Jennifer Cuillerier Arens, hereby grant a non-exclusive, for the full term of copyright protection, license to my University, The University of Prince Edward Island:

(a) to archive, preserve, produce, reproduce, publish, communicate, convert into any format, and to make available in print or online by telecommunication to the public for non-commercial purposes;
(b) to sub-license to Library and Archives Canada any of the acts mentioned in paragraph (a).

I undertake to submit my thesis/dissertation, through my University, to Library and Archives Canada. Any abstract submitted with the thesis/dissertation will be considered to form part of the thesis/dissertation.

I represent that my thesis/dissertation is my original work, does not infringe any rights of others, including privacy rights, and that I have the right to make the grant conferred by this non-exclusive license.

If third party copyrighted material was included in my thesis/dissertation for which, under the terms of the *Copyright Act*, written permission from the copyright owners is required I have obtained such permission from the copyright owners to do the acts mentioned in paragraph (a) above for the full term of copyright protection

I retain copyright ownership and moral rights in my thesis/dissertation, and may deal with the copyright in my thesis/dissertation, in any way consistent with rights granted by me to my University in this non-exclusive license.

I further promise to inform any person to whom I may hereafter assign or license my copyright in my thesis/dissertation of the rights granted by me to my University in this non-exclusive license.

Signature	**Date**

University of Prince Edward Island

Faculty of Veterinary Medicine

Charlottetown

CERTIFICATION OF THESIS WORK

We, the undersigned, certify that Jennifer Cuillerier Arens, candidate for the degree of Doctor of Philosophy, has presented her/his thesis with the following title:

Heterologous production of bioactive terpenes and discovery of novel natural products from members of the genus *Kitasatospora*

that the thesis is acceptable in form and content, and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate through an oral examination held on May 25th 2015

Examiners

Dr. David Jakeman (External)

Dr. Tarek Saleh (Chair)

Dr. Russell Kerr (Supervisor)

Dr. Spencer Greenwood

Dr. Christopher Kirby

ABSTRACT

Nature has provided scientists with numerous bioactive molecules with great potential as therapeutic agents for a range of human and animal diseases. Exploring terrestrial and marine environments for macro- and microorganisms capable of producing these valuable secondary metabolites remains an important source of new drug-leads. Unfortunately, the development of many natural products into full-fledged pharmaceutical products is frequently hindered by the inability of the natural source to provide sufficient quantities to support clinical evaluation.

To address the supply issue hindering the development of many terpenes, a heterologous expression platform capable of high-level production of bioactive terpenes was developed. The metabolic profile of *Streptomyces lividans* TK24 was first improved by genetically disrupting the production of actinorhodin and undecylprodigiosin via markerless deletions. This removed any interference from these metabolites during biological and chemical assays and eliminated competition for common biosynthetic precursors. Terpentecin biosynthetic genes were then expressed in the genetically modified *S. lividans* host under the control of a constitutive promoter, though no detectable yield of the diterpene were present in the heterologous host.

This work led to the identification of two new glucosylated polyketides, satosporin A and B, isolated from the strain *Kitasatospora griseola* MF730-N6. Structure elucidation and stereochemical assignments were performed using spectroscopic analyses, computational methods and chemical derivatizations. Satosporins possess an unprecedented tricyclic ring system comprising an oxo-decalin unit fused to an 8-membered lactone and 11 chiral centers. The aglycone, satosporin C, was selectively obtained by enzymatic deglucosylation of satosporin A using β-glucosidase from almonds.

The satosporing biosynthetic gene cluster was identified from the *K. griseola* genome via a combination of degenerate primers and Illumina® and Pacific Biosciences® sequencing technologies. The DNA sequences were analyzed by antiSMASH and BLAST® and revealed a 67 kb region that putatively encoded a cytochrome P450 monoxygenase, two large multimodular polyketide synthases, a unique fused ketosynthase-cytochrome P450 monoxygenase, a glycosyltransferase as well as regulatory and self-resistance genes. A knock-out experiment within module 1 of a polyketide synthase (*satB*) confirmed the involvement of this gene cluster in satosporin biosynthesis as production ceased following gene disruption.

The Pacific Biosciences® sequencing data along with closure of gaps by sequencing overlapping amplicons led to the draft genome of *K. griseola* in 8 contigs totalling 7,966,157 bp and an overall G + C content of 72.7%. Annotation using RAST revealed 9 complete + 5 incomplete rRNA operons and 7,226 protein encoding sequences. The *K. griseola* genome was similar to other kitasatosporae and streptomycetes in terms of developmental and regulatory genes and antiSMASH analysis revealed 23 secondary metabolite biosynthetic gene clusters, 15 of which had unknown products.

Over 75% of the bioactive metabolites produced by actinomycete bacteria are from members of the genus *Streptomyces*. There is, however, a growing interest in bioprospecting from “rare” or non-*streptomyces* actinomycetes such as kitasatosporae. Twelve kitasatosporae were thus cultured in two media and metabolomic analysis using principal component analysis was used to rapidly identify putatively new natural products. Through collaborative work with MMS student Krista Gill, cystargamide and cystargolide A and B were identified as new compounds.

ACKNOWLEDGEMENTS

I would like to start by thanking Dr Russell Kerr for having taken me on as a graduate student back in 2009. Throughout the years, you have given me a great deal of flexibility and independence and you have entertained many reseach ideas that, although unplanned, have yielded great results; without these leaps of faith, I would not have learned as much about so many research areas as I have while under your supervision. To my research manager, Bradley Haltli, I thank you for sharing your knowledge and skills and always being available to troubleshoot and brainstorm with me. Your passion and enthusiasm for quality scientific research has helped shape me into the scientist I am today and has, in the process, made me fall in love with bacterial genetics and molecular biology.

We are very fortunate in the Kerr laboratory to have access to excellent research managers with distinct areas of expertise and a strong desire to help and teach. To Fabrice Berrué, I thank you for being patient with me as I learned numerous new skills related to chemistry. You were always available when I had questions and always made sure I understood the concept fully before sending me off, even if that meant you had to repeat yourself a few times. Thanks to David Overy and Martin Lanteigne for performing the antimicrobial assays for our lab and being available to answer any questions related to them as well. Thanks to all the lab members who, thoughout the years, have become great friends. To my officemates and pseudo-officemates (Becca, Veronica, Erin, Stacey, Alyssa, Ghada, and Amanda), I have become especially close to you and want to thank you for all the talks we have had over the year, science related or not and for all the fun times we shared together.

I would like to acknowledge my current supervisory committee (Dr Russ Kerr, Dr Spencer Greenwood, Dr John Burka, Dr Rick Cawthorn and Dr Jason McCallum) as well as past members (Dr Byeong Hwa Jeon and Dr Junzen Zhang) for showing interest in my research and providing valuable feedback throughout my degree and regarding this thesis.

This research would not have been possible if not for financial support (stipend and research funds) from the Atlantic Veterinary College and Department of Biomedical Sciences (graduate student scholarship and other awards/grants), the Natural Sciences and Engineering Reserch Council of Canada, the Atlantic Innovation Fund, the Canada Research Chair Program and the Lévesque Foundation.

Furthermore, I would like to acknowledge Tricia Boland for running LCMS experiments, Dr Nathan Magarvey for the Illumina® sequencing data, Genome Québec for PacBio sequencing services, Dr Alison Thompson for providing circular dichroism spectra and Dr Jason Peason for collaborating with us and sharing his computational chemistry expertise. A sincere thank you to the staff in biomedical sciences, Debbie Gallant, Suzette Acorn and Sherri Pineau, for always being eager to help and for the cheerfulness you bring to the department.

I would like to thank my parents for shaping me into the person I have become and allowing and supporting me in the pursuit of my dreams eventhough they have taken me further away from home. Thank you for believing in me; I strive to make you proud. To my brother, thank you for keeping me spontaneous and teaching me that winging it every once in a while is good for me. Thanks to all of my extended family for always being interested as to what I am working on and thanks for traveling to PEI so I could share the wonders of this amazing island with you. Thank you to my mother-in-law for raising a wonderful man, for welcoming me into your life and for supporting us throughout our studies.

To my close friends from childhood, high school, college and university, thank you for your support and encouragements throughout my seemingly never ending studies. No matter the distance, I forever cherish our friendships.

Finally, to my husband, how can I put in words how grateful I am for everything you give me... You are my best friend, my support system, my sidekick and my daily source of love, laughter and aww. Thank you for marrying me.

To my family…

…the one I was fortunate to be born into and the one I chose to marry into.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	absorbance
A	adenylation
Å	angstrom
ABC	ATP-binding cassette
ACN	acetonitrile
ACP	acyl-carrier protein
ACT	actinorhodin
ANOVA	analysis of variance
antiSMASH	antibiotic and secondary metabolites analysis shell
ARS	Agriculture Research Service
AT	acyltransferase
BLAST®	Basic Local Alignment Search Tool
bp	base pair
C	condensation
C_{18}	octadecylsilane
CCS	circular consensus sequence
CD_3OD	deuterated methanol
cDNA	complementary deoxyribonucleic acid
CDS	coding sequences
CLR	continuous long read
CoA	coenzyme A
COSY	correlation spectroscopy
DAP	diaminopimelic acid
DCM	dichloromethane
DH	dehydratase
Dha	dehydroalanine
Dhb	dehydrobutyrine
DMAP	4-dimethylaminopyridine
DMAPP	dimethylallyl diphosphate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DXP	1-deoxy-D-xylulose 5-phosphate
ε	extinction coefficients
ECD	electronic circular dichroism
ED_{50}	median effective dose
ELSD	evaporative light scattering detector
eq	equivalents
ER	enoyl reductase
ESI	electrospray ionization
EtOAc	ethyl acetate
FPP	farnesyl diphosphase
gDNA	genomic deoxyribonucleic acid
GGPP	geranylgeranyl diphosphate

GPP	geranyl diphosphate
HGAP2	Hierarchical Genome Assembly Process 2
HMBC	heteronuclear multiple bond correlation
HMG-CoA	3-hydroxy-3-methyl-glutaryl-CoA
HPLC	high performance liquid chromatography
HRESIMS	high resolution electrospray ionization mass spectrometry
HRMS	high resolution mass spectrometry
HSQC	heteronuclear single quantum coherence
l	path length
IC_{50}	half maximal inhibitory concentration
IPA	isopropyl alcohol
IPOD	International Patent Organism Depositary
IPP	isopentenyl diphosphate
ISP3	international *Streptomyces* project medium 3
J	coupling constant
kb	kilobase pair
KO	knock-out
KR	ketoreductase
KS	ketosynthase
LB	Luria-Bertani
LC	liquid chromatography
LC-HRMS	liquid chromatography-high resolution mass spectrometry
m/z	mass to charge ratio
Mb	megabase pair
MeOH	methanol
MEP	2-*C*-methyl-D-erythritol 4-phosphate
MEV	mevalonate
MIBiG	minimal information about a biosynthetic gene cluster
MIC	minimum inhibitory concentration
MMS	molecular and macromolecular sciences
MPA	methoxyphenylacetic acid
MRSA	methicillin-resistant *Staphylococcus aureus*
MS	mass spectrometry
MS^n	tandem mass spectrometry
MYM	maltose-yeast extract-malt extract
N50	minimum fragment length to generate half of the genome assembly
nanoDESI	nanospray desorption ionization
NCBI	national center for biotechnology information
NEB	New England Biolabs
NMPDR	National Microbial Pathogen Data Resource
NMR	nuclear magnetic resonance
NOE	nuclear overhauser effect
NOESY	nuclear overhauser effect spectroscopy
NRP	nonribosomal peptide

NRPS	nonribosomal peptide synthase
nt	nucleotide
ORF	open reading frame
OSMAC	one strain/many compounds
PacBio	Pacific Biosciences®
PCA	principal component analysis
PCM	polarizable continuum solvent model
PCP	peptidyl carrier protein
PCR	polymerase chain reaction
PDA	photodiode array detector
pDNA	plasmid deoxyribonucleic acid
PEG	protein coding sequence
PK	polyketide
PKS	polyketide synthase
R	*rectus*
RAST	rapid annotation using subsystem technology
RBS	ribosome binding site
RED	undecylprodigiosin
RiPPs	ribosomally synthesized and post-translationally modified peptides
RNA	ribonucleic acid
ROESY	rotating frame overhauser effect spectroscopy
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
RT-PCR	reverse transcription-polymerase chain reaction
S	*sinister*
SALP	SsgA-like protein
SBSPKS	structure based sequence analysis of polyketide synthases
SMRT®	single molecule real-time
SPE	solid phase extraction
TBME	*tert*-butyl methyl ether
TCA	tricarboxilic acid
TDDFT	time-dependent density functional theory
TFA	trifluoroacetic acid
t_R	retention time
tRNA	transfer ribonucleic acid
TSB	tryptic soy broth
UV	ultraviolet
ν_{max}	wavenumber absorption maxima
VRE	vancomycin-resistant *Enterococcus*
WT	wild-type
YEME	yeast extract-malt extract
ZMW	zero-mode waveguide

CHAPTER 1 – AN INTRODUCTION TO NATURAL PRODUCTS- CURRENT CHALLENGES AND THESIS RESEARCH GOALS

1.1 Natural products and their biosynthesis

Natural products, or secondary metabolites, are non-essential compounds of tremendous structural diversity biosynthesized by various organisms that are believed to provide a selective advantage to the producer[1, 2]. The compounds fall into the following classes based on their biosynthetic origin and structural composition: terpenes, polyketides, nonribosomal peptides, alkaloids, shikimic acid derivative and others.

Terpene (or isoprenoid) natural products include compounds such as artemisinin, geosmin, paclitaxel, terpentecin, eleutherobin and pseudopterosins. They are biosynthesized from consecutive head-to-tail condensation of two universal isoprene precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), to give rise to mono- (C_{10}), sesqui- (C_{15}) and diterpenes (C_{20}) via geranyl diphosphate (GPP), farnesyl diphosphase (FPP) and geranylgeranyl diphosphate (GGPP) linear intermediates[3]. IPP and DMAPP are produced via one of two biosynthetic routes, the mevalonate (MEV) and the non-mevalonate pathways. The mevalonate pathway is present in eukaryotes including all mammals, the cytosol and mitochondria of plants and fungi, archae and a few bacteria. It converts 3 molecules of acetyl-coenzyme A (CoA) into IPP via a mevalonate intermediate[4] (Figure 1.1). In contrast, the non-mevalonate pathway (also referred to as the 2-*C*-methyl-D-erythritol 4-phosphate (MEP) and 1-deoxy-D-xylulose 5-phosphate (DXP) pathways) is found mainly in bacteria, cyanobacteria, green algae, and in the plastid of plants and converts pyruvate and glyceraldehyde-3-phosphate into IPP and DMAPP via a methylerythritol phosphate intermediate[4] (Figure 1.1). The great structural diversity of terpenes is then achieved by reactions such as cyclization, methylation, hydroxylation, epoxidation and glycosylation[4]. Actinomycetes such as *Kitasatospora griseola* MF730-N6, *Streptomyces* sp. CL 190, *Streptomyces* sp. KO-3988 and *Actinoplanes* sp. A40644 have been reported to utilize both pathways, where mevalonate biosynthetic genes were usually clustered with other terpene biosynthetic genes[5].

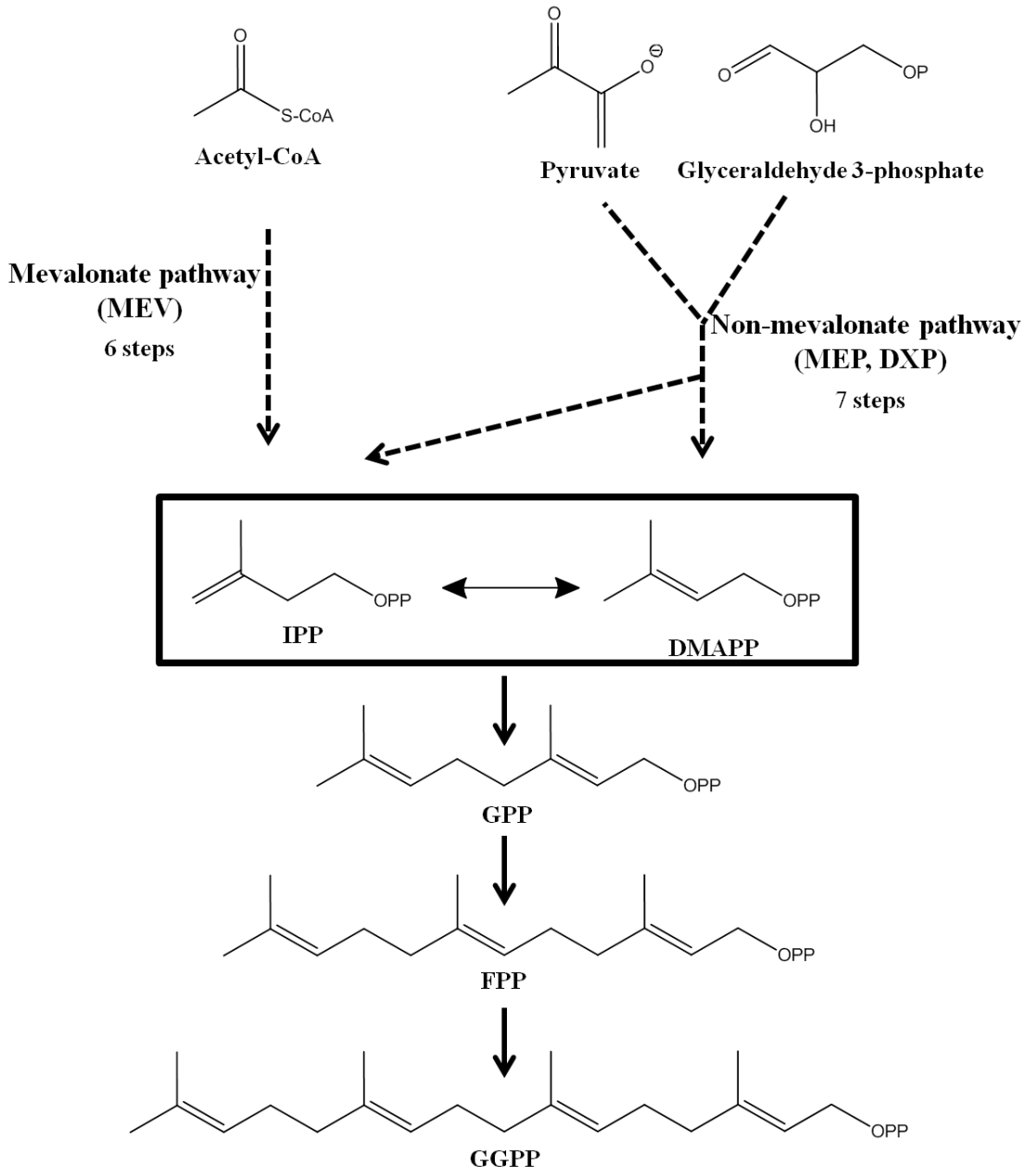


Figure 1.1. Two biosynthetic routes, the mevalonate and non-mevalonate pathways, leading to the production of the two isoprene precursor units which are then condensed by prenyltransferases to form the basic skeleton of all terpenes; IPP: isopentenyl diphosphate, DMAPP: dimethylallyl diphosphate, GPP: geranyl diphosphate, FPP: farnesyl diphosphase, GGPP: geranylgeranyl diphosphate. Reproduced, with permission, from Kirby and Keasling, 2008[6].

Polyketide natural products include compounds such as erythromycin, rapamycin, lovastatin and actinorhodin that are biosynthesized via the condensation of a starter unit (usually acetyl-CoA and propionyl-CoA) with extender units such as malonyl-CoA and methylmalonyl-CoA[7]. The enzymes involved in polyketide assembly are polyketide synthases (PKSs) and are generally categorized as a type I if it is a large multimodular polypeptide containing modules that are each responsible for a single chain extension step as in erythromycin biosynthesis, or as type II if the group of polypeptides are reused for each chain elongation as is the case in actinorhodin biosynthesis[8]. During polyketide biosynthesis, the extender unit is selected and tethered to the acyl-carrier protein (ACP) by the acyltransferase (AT) following which it is condensed with the previous ketide by the ketosynthase (KS), resulting in a two-carbon extension of the ketide. The presence of one or more of the β-processing enzymes ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) will reduce the β-carbonyl group to an alcohol, alkene or alkane, respectively; along with post-PKS tailoring enzymes, they contribute to the incredible structural diversity observed in polyketides[9].

Nonribosomal peptides are biosynthesized similarly to polyketides, though using amino acids as building blocks and nonribosomal peptide synthases (NRPSs) and adenylation (A), condensation (C) and peptidyl carrier protein (PCP) as the core domains and epimerases, cyclases, oxidases and methyltransferases as tailoring enzymes[10]. Alkaloids are nitrogen containing natural products often biosynthesized from ornithine, lysine, tyrosine and tryptophan, but where the nitrogen atom may also arise from other sources[11]. Shikimic acid derivatives, as the name implies, are compounds derived from shikimic acid and can lead to phenolic natural products[11]. Other natural products include ribosomally synthesized and post-translationally modified peptides (RiPPs)[12].

Secondary metabolites are produced by diverse organisms such as bacteria, fungi, plants, and invertebrates of both terrestrial and marine habitats and can provide functional roles in defense (and offense) and communication[1, 13]; the production of secondary metabolites can also

play important roles in symbiotic relationships[13]. It is becoming increasingly evident that bacterial symbionts can be responsible for the production of secondary metabolites that were initially thought to be produced by invertebrates. Such has been the case for bryostatin, onnamide/peredin, psymberin, patellamides, cyanobactins and ET-743[14-16]. Though many more natural products are suspected to be produced by a microbial symbiont, challenges in culturing these symbionts have made it difficult to unequivocally identify the true producers and instead, genetic techniques must be used to provide supporting evidence for microbial biosynthesis[17].

Although the link between the natural product and its ecological importance is often unknown, numerous compounds have biological properties that can be exploited by humans[1]. We have indeed benefited from natural products for thousands of years, beginning with the use of ethnobotanical remedies[18] and more recently with the isolation of pure bioactive compounds by the pharmaceutical industry for use as antimicrobial, anticancer, antiviral, anti-inflammatory and anti-hypertensive agents[19, 20]. Natural product drug discovery is a lengthy and costly process which includes the discovery of novel compounds with promising biological activity, followed by preclinical and clinical trials (Phase I, II and III)[21]. Although challenges may be encountered at every step, focus on challenges pertaining to the early steps of the process will be discussed here as this constitutes the focus of the research within the Kerr laboratory: to identify new natural products with potential uses as human and animal therapeutics. The main challenges faced are the high occurrence of compound rediscovery (including the difficulty of discovering novel natural products) and the inability to provide required amounts of the natural product, also referred to as the supply issue[22].

1.2 Strategies to discover novel natural products

To discover new secondary metabolites, one must first be able to rapidly identify which compounds in an extract are known, a process known as chemical dereplication. This important step ensures minimal time and resources are allocated to the purification and structure

elucidation of known compounds. Numerous strategies have been developed to enable chemical dereplication and usually rely on a combination of chromatographic methods (such as liquid chromatography (LC) and solid phase extraction (SPE)) coupled with spectrometric and spectroscopic techniques (such as mass spectrometry (MS) and nuclear magnetic resonance (NMR)) along with searchable databases (such as AntiBase, Marinlit, AntiMarin and the Dictionary of Natural Products)[23]. While advances in analytical techniques have allowed for reduced preparation times and higher instrument sensitivity (tandem mass spectrometry (MS^n), nanospray desorption ionization (nanoDESI) MS, high resolution MS, two-dimentional NMR)[22], advances in computerized methodologies have enabled the organization of the plethora of information in a meaningful way[24, 25]; together, they have enabled the efficient dereplication of large natural product extract libraries. In addition to the rapid identification of known compounds, metabolic profiling of natural product extracts can also be used to dereplicate bacterial samples in order to compile an organism library based on unique metabolic profiles rather than 16S ribosomal ribonucleic acid (rRNA)-based phylogenetic analysis[24].

In addition to discarding known metabolites, researchers can improve the likelihood of discovering new natural products by collecting organisms from unique environments and applying selection methods to remove common species and/or select for unique species within the community assemblage as commonly isolated organisms will likely lead to the isolation of known compounds[26-28].

1.3 Sequencing reveals numerous cryptic biosynthetic pathways within bacterial genomes

Even when care is taken to study underexplored organisms from unique habitats, natural product discovery is hindered by the organism's secondary metabolite regulatory network and the inability of culturing conditions to trigger the production of numerous biosynthetic pathways. It is now apparent that numerous microorganisms, especially actinomycetes, have the potential to produce many more natural products than reported using traditional fermentation and

extraction techniques[29]. A well understood example is that of the model organism *Streptomyces coelicolor* (Figure 1.2), which was believed to produce only a handful of compounds until its genome was sequenced in 2002, revealing over 20 secondary metabolite biosynthetic gene clusters[30]; as a result of genome mining, 10 additional compounds have now been described from this organism[31].

Strategies to access these cryptic natural products include both fermentation techniques and genetic manipulations to specifically or globally enhance secondary metabolite production[31]. Manipulating culture conditions such as media composition, nutrient availability, temperature, osmotic pressure and exposure to antibiotics, small molecules and other organisms to trigger an increase in production using the one strain/many compounds (OSMAC) approach has proven successful but results tend to be unpredictable and are difficult to interpret genetically[29]. On the other hand, genetic approaches are more tractable but are more laborious; they include the activation/overexpression of global or pathway specific regulators, inactivation of repressors, improvements to precursor availability by metabolic engineering and finally heterologous expression of the natural product[31].

A recent example of a formerly cryptic natural product whose production was triggered due to variation in culture condition is that of coelichelin in *S. coelicolor*. Since genome mining revealed that a putative siderophore (encoded by an NRPS) would sequester iron due to the presence of hydroxamic acid groups; fermentation of *S. coelicolor* in low iron conditions led to detectable yields of the new compound coelichelin[32]. An example of using genetic manipulations to activate cryptic genes has been exemplified by kitasetaline, where genome mining of *Kitasatospora setae* revealed paralogs of a γ-butyrolactone autoregulator which had been previously shown to regulate secondary metabolite production[33]; the inactivation of one of the paralog, *ksbC*, led to detectable levels of the new compound kitasetaline[34].

With the reduced costs and improved capabilities of sequencing technologies (Illumina®, Ion Torrent™, Pacific Biosciences® (PacBio) and nanopore sequencing[35-37]), as well as

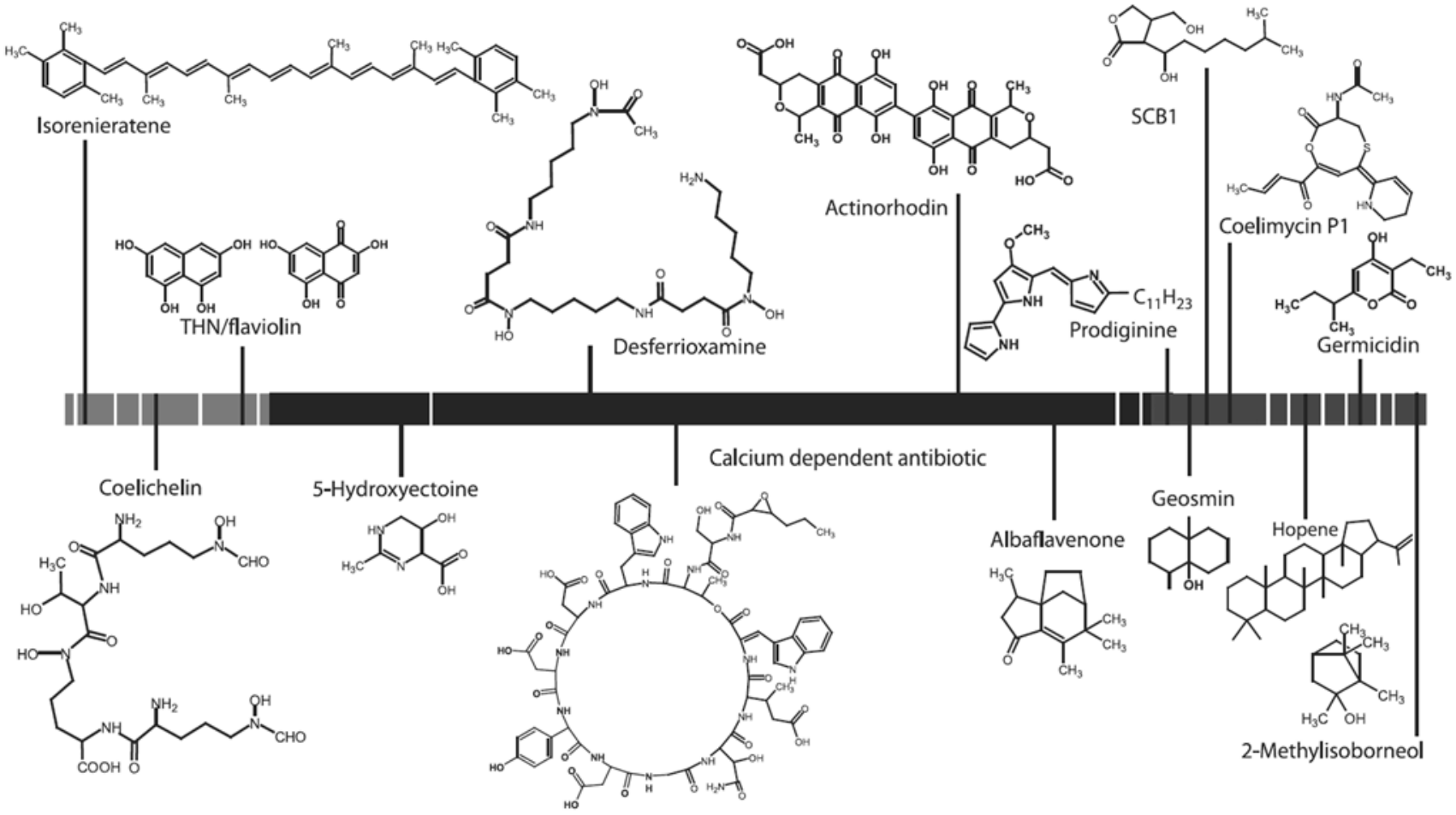


Figure 1.2. Secondary metabolites biosynthetic gene clusters along the *S. coelicolor* genome; gene clusters of known metabolites are shown in black while those with unknown products are represented by a white vertical line. Reproduced, with permission, from Craney et al.[31]

improvements in genome annotation and secondary metabolite gene prediction software such as the antibiotic and secondary metabolites analysis shell (antiSMASH)[38] and eventually the minimal information about a biosynthetic gene cluster (MIBiG, unpublished), genome mining is sure to continue to provide numerous opportunities to enhance production of low-yield or cryptic natural products from various producers.

1.4 The supply issue can limit the development of natural products into commercial products

As mentioned previously, natural products can sometimes be difficult to purify due to their cryptic nature and low yields. In order to progress through the drug development pipeline, large amounts (~100 g) of the compounds are necessary to begin clinical trials[19]. Since natural products usually comprise less than 1% by weight of the crude extract[22], the supply issue is a common obstacle in drug discovery programs, especially when the source of the natural product cannot be readily cultured[39].

Pseudopterosins and eleutherobin are two glycosylated diterpenes isolated from corals of the genera *Antillogorgia* (previously *Pseudopterogorgia*) and *Eleutherobia*, with promising anti-inflammatory[40] and antitumor activities[41], respectively. Even though pseudopterosins constitute up to 5% of the coral crude extract[42], their supply for the cosmetic industry and clinical trials continues to be a challenge[43]. Similarly, the very low yields of eleutherobin (0.01%) in crude extracts of the coral[41] has prevented even preclinical evaluations of the compound as an anticancer drug with similar potency to the currently used drug Taxol®[44, 45].

Strategies to overcome the supply issue include large scale collections of the producing organism, total and semisyntheses, searching for an alternative producer of bacterial origin, metabolic engineering or random mutagenesis of the producing organism to eliminate competing pathways and bottlenecks and finally expression of the biosynthetic genes in a heterologous host[46, 47]. Though the Kerr laboratory investigates a number of these strategies, a portion of the

following thesis will focus on the use of heterologous expression to achieve sustainable production of natural products.

1.5 The use of heterologous hosts to produce important natural products

Heterologous expression involves the incorporation of appropriate biosynthetic genes from the producing organism into a surrogate (heterologous) host that does not normally produce the natural product of interest[48]. Parameters to consider when choosing the heterologous host include ease of culturing (rapid growth in inexpensive medium), well developed and tractable genetic methods for incorporating and expressing foreign deoxyribonucleic acid (DNA), the presence of the necessary precursors and cofactors and finally a comprehensive understanding of the organism's regulatory networks and secondary metabolite production (both isolated and cryptic)[49].

The use of heterologous expression is a valuable tool in characterizing natural product biosynthetic pathways[49] and is often used to express a newly elucidated biosynthetic gene cluster as additional evidence that the genes reported are in fact responsible for the biosynthesis of the compound[50-55]. Most heterologous expression experiments to date have used *Escherichia coli* and *Saccharomyces cerevisiae* as hosts due to their rapid growth rates and the availability of molecular techniques developed specifically for these organisms[48]. For experiments where the objective is to characterize biosynthetic pathways and only detectable levels of the natural product are required, these model organisms may serve as ideal hosts. However, if the purpose is to have an economic and renewable source of a valuable natural product, alternative hosts may provide superior yields without requiring decades of pathway engineering and fermentation optimization.

Although the development of a single host for the high-level production of all natural product classes is unlikely, class-specific heterologous hosts hold potential due to the common biosynthetic origins of compounds within a class[49]. Since all terpenes share the same the two

universal precursors (IPP and DMAPP), a heterologous host capable of producing the two compounds in elevated amounts holds great promise for producing an array of terpenes in high yield.

Previous work attempting to optimize the production of high-value, complex terpenes in heterologous hosts has focused mainly on artemisinin and paclitaxel in *E. coli* and *S. cerevisiae* hosts. Successful yield increases have been achieved via an increase in copy number of genes involved in rate limiting steps within the MEV and MEP pathways, especially when combined with a high flux towards downstream steps as some accumulating intermediates were proven to be toxic[56, 57]. The use of exogenous, engineered and codon-optimized genes have provided enzymes with increased activity and, along with eliminating competing pathways[58], have also resulted in increased terpene yields[56, 59]. Combining these engineering steps with fermentation, extraction and processing optimization have, for example, finally resulted in yields of 25 g/L of artemisinic acid, the target yield for economic production of semisynthetic artemisinin[60].

Although sufficient yields of heterologous artemisinic acid were eventually obtained, decades of contribution from numerous research groups were necessary to achieve these results. An alternative approach to this lengthy process involves the use of mutagenic agents to generate the desired genotype via random mutations, a process that has been used since the 1940s to provide overproducing strains to the pharmaceutical industry[61].

Successful random mutagenesis as an approach to increase terpene titers is not conducive to high-throughput screening as it relies on lengthy chemical extractions and expensive analytical techniques. Given the low probability of obtaining the desired phenotype, numerous rounds of mutagenesis and thousands, if not hundreds of thousands, of colonies must be screened[62]; a rapid method to identify terpene overproducing mutants is thus essential. The use of growth inhibition zones in a soft-agar overlay assay to estimate the production of the antibacterial compound has been used in traditional strain improvement programs[61] and, thus, its

use as a rapid terpene screening method would permit the generation of a new platform host for the high-level production of terpenes via random mutagenesis.

Streptomycetes represent promising hosts due to their innate ability to produce numerous natural products of various classes including terpenes; the appropriate precursors and cofactors required for terpene biosynthesis may thus already be present[63]. Streptomycete hosts have previously been used for the heterologous production of bacterial terpenes such as platencin, phenalinolactone, terpentecin, cyclooctadin and pentalenolactone[50, 53, 55, 64, 65] for the purpose of pathway characterization, demonstrating that streptomycetes make adequate hosts for the production of bacterial terpenes.

As there is a need for inexpensive, environmentally friendly and sustainable terpene production methods, a portion of this thesis will focus on the development of a streptomycete heterologous host to circumvent the supply issue as it pertains to the production of terpenes specifically.

1.6 Thesis research

The general objective within the Kerr laboratory is to discover new natural products for applications in human and animal wellness. Members within the lab employ a variety of strategies to achieve this goal such as exploring bacterial and fungal communities in various environments, using molecular techniques to identify potential producers of secondary metabolites, developing new analogues of known compounds via synthetic routes, and using metabolomic analyses to identify potentially new natural products.

One primary objective of my thesis work was to use molecular engineering and strain improvement methods to develop a new streptomycete heterologous host that could be used to provide a more efficient production of known natural products (Chapter 2). As actinomycetes are known to be prolific producers of natural products, special attention was given to the extracts from such organisms used during this research and as a result, a new natural product (satosporin) was isolated and characterized (Chapter 3). Further examination into the biosynthesis of this new

compound was accomplished by sequencing the genome of the producing organism, allowing for the description of the satosporin biosynthetic gene cluster (Chapter 4). Simultaneously, this data also allowed genomic comparisons to be made with the few kitasatosporae genomes previously available and the assessment of secondary metabolite production potential of the organism (Chapter 5). Finally, a study was undertaken to compare the metabolomes of numerous kitasatosporae and principal component analysis was used to rapidly identify potentially novel natural products (Chapter 6).

The specific thesis objectives are:

Objective 1: To develop a general platform host for the production of bioactive terpenes.

Objective 2: To discover new natural products from actinomycetes of the genus *Kitasatospora*.

Objective 3: To investigate the genetic origin of satosporins.

Objective 4: To describe the genome of *Kitasatospora griseola*.

CHAPTER 2-HETEROLOGOUS EXPRESSION OF TERPENTECIN IN A METABOLICALLY SIMPLIFIED STRAIN OF *STREMPTOMYCES LIVIDANS*

2.1 INTRODUCTION

Terpene natural products exhibit promising antimicrobial, antiviral, anti-hypertensive, anti-inflammatory, antiparasitic and anticancer activities and have been isolated from a variety of organisms including corals, sponges, algae, bacteria, plants and fungi from both terrestrial and marine environments[3]. Their development into commercially available pharmaceutical products is often hindered by insufficient yields produced by the native host, commonly referred to as the supply issue. Eleutherobin, for example, is a promising diterpene with potent anticancer activity but represents only 0.01% of the crude extracts of the coral *Eleutherobia* sp.[41] Due to the structural and stereochemical complexities of these compounds, chemical synthesis rarely provides a financially feasible alternative production method and large-scale harvest of terpene producing marine invertebrates can have devastating effects on marine ecosystems such as coral reefs[66]. Consequently, there is pressing need for inexpensive, environmentally friendly and sustainable terpene production methods.

Heterologous host expression provides such a solution. In this approach, a well characterized and genetically malleable microorganism is used as an alternative host to produce these bioactive natural products via the incorporation of the terpene biosynthetic genes from the native producer. Since all terpenes are biosynthesized from the same two precursors: IPP and DMAPP, the development of a universal host capable of producing these two metabolites in high-yields is an attractive proposal towards achieving high yields of various bioactive terpenes. Heterologous hosts such as *Escherichia coli*, *Saccharomyces cerevisiae*, *Streptomyces avermitilis*, *Streptomyces coelicolor* and *Streptomyces lividans* have been used for terpene production with limited success[63]. Recently, however, heterologous amorphadiene production in *S. cerevisiae* has finally reached respectable yields, following 11 years of optimization, where the necessary artemisinic acid yield of 25 g/L to supply the world with the antimalarial drug artemisinin was achieved in 2013[58]. This long and laborious metabolic engineering process could be circumvented by the use of the traditionally used random mutagenesis, where random

mutation(s) are generated by iterative exposures to mutagenic chemicals or ultraviolet (UV) radiation. This method has been extensively employed in strain development of antibiotic producing microorganisms since the discovery of penicillin, where production yields were improved in *Penicillium chrysogenum* from 60 mg/L to 1.8 g/L[62]. Although successful, random mutagenesis as an approach to increase terpene titers is not conducive to high-throughput screening and given that the probability of identifying improved mutants varies between one in 1000 to one in 100,000[62], a rapid method to identify terpene overproducing mutants is essential. Since antibiotic production can be rapidly estimated based on growth inhibition zones in an agar based assay, the production of an antibacterial terpene in the heterologous host would allow the detection of high terpene producing mutants. Terpentecin is a diterpene isolated from the actinomycete *Kitasatospora griseola* MF730-N6 (previously *Kitasatosporia griseola* and *Streptomyces griseolosporeus*) from soils of Nara Prefecture, Japan[67]. The compound showed antimicrobial activity against Gram-positive and Gram-negative bacteria including *Staphylococcus aureus* (minimum inhibitory concentration (MIC) <0.05 µg/mL) and *Bacillus subtilis* NRRL B-558 (MIC <0.05 µg/mL)[67] and the gene cluster responsible for terpentecin biosynthesis in *K. griseola* was elucidated in 2001[64]. The production of terpentecin in the heterologous host would thus allow a rapid antibacterial agar-based assay to track increases in terpene production following each round of mutagenesis.

The objective of this research was to develop a strain of *Streptomyces lividans* TK24 capable of producing high yields of terpenes. To this end, the strain was first genetically modified to eliminate the production of two native secondary metabolites: actinorhodin and undecylprodigionsin. This eliminated any interference from these metabolites during biological and chemical assays and eliminated competition for common biosynthetic precursors. Diterpene biosynthetic genes responsible for the production of the antibacterial natural product terpentecin were then cloned from *Kitasatospora griseola* and expressed into the newly developed host under the control of a constitutive promoter using a newly developed vector and its production in

the heterologous host was assessed. An overview of the strategy to heterologously produce terpentecin is graphically represented in Figure 2.1.

2.2 MATERIALS AND METHODS

Restriction enzymes, T4 DNA ligase, Antarctic Phosphatase and competent *Escherichia coli* 5α and 10β were purchased from New England Biolabs (NEB, Whitby, ON). KOD Hot Start DNA Polymerase (TOYOBO), EconoTaq® PLUS GREEN 2X Master Mix (Lucigen) and E.Z.N.A.® plasmid mini kit (Omega Bio-Tek) were from VWR (Missassauga, ON). Chemical reagents were from Sigma-Aldrich (Oakville, ON). Gel electrophoresis was performed using 1% agarose, TAE buffer and 0.05 μg/mL ethidium bromide. Column and gel purifications were performed using EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic) from VWR. General procedures were performed according to Sambrook et al.[68] Chemically competent cells were prepared following the rubidium chloride method (published under NEB protocols and modified by Joseph Utermohlen). A list of plasmids and primers used and constructed is listed in Table 2.1 and Table 2.2 and illustrated in Figure 2.2. Media recipes are listed in Table 2.3.

Liquid chromatography-high resolution mass spectrometry (LC-HRMS) analysis was performed at various steps using Thermo Scientific™ (Waltham, USA) LTQ Exactive™ equipment with a Core Shell Kinetex® 1.7 μm C_{18}, 50 X 2.1 mm and 100 Å column, photodiode array detector (PDA), SEDEX 80 Sedere LT-ELSD evaporative light scattering detector (ELSD) and a Finnigan LXQ ion trap mass spectrometer in positive mode equipped with an electrospray ionization source (ESI) detecting *m*/*z* between 190 and 2000. Chromatographic separation was achieved with a 4.8 min gradient of 95:5 water (H_2O):acetonitrile (ACN) to 100% ACN, 3.2 min of 100% ACN and finally a return to starting conditions over 3 min; solvents contained 0.1% formic acid and a flow rate of 500 μL/min was used.

2.2.1 Construction of *S. lividans* TK24 ΔactΔred

To eliminate the production of actinorhodin (ACT) (**1**) and undecylprodigiosin (RED) (**2**) (Figure 2.3 (A)) in *S. lividans* TK24, a portion of the act and the entire red biosynthetic gene

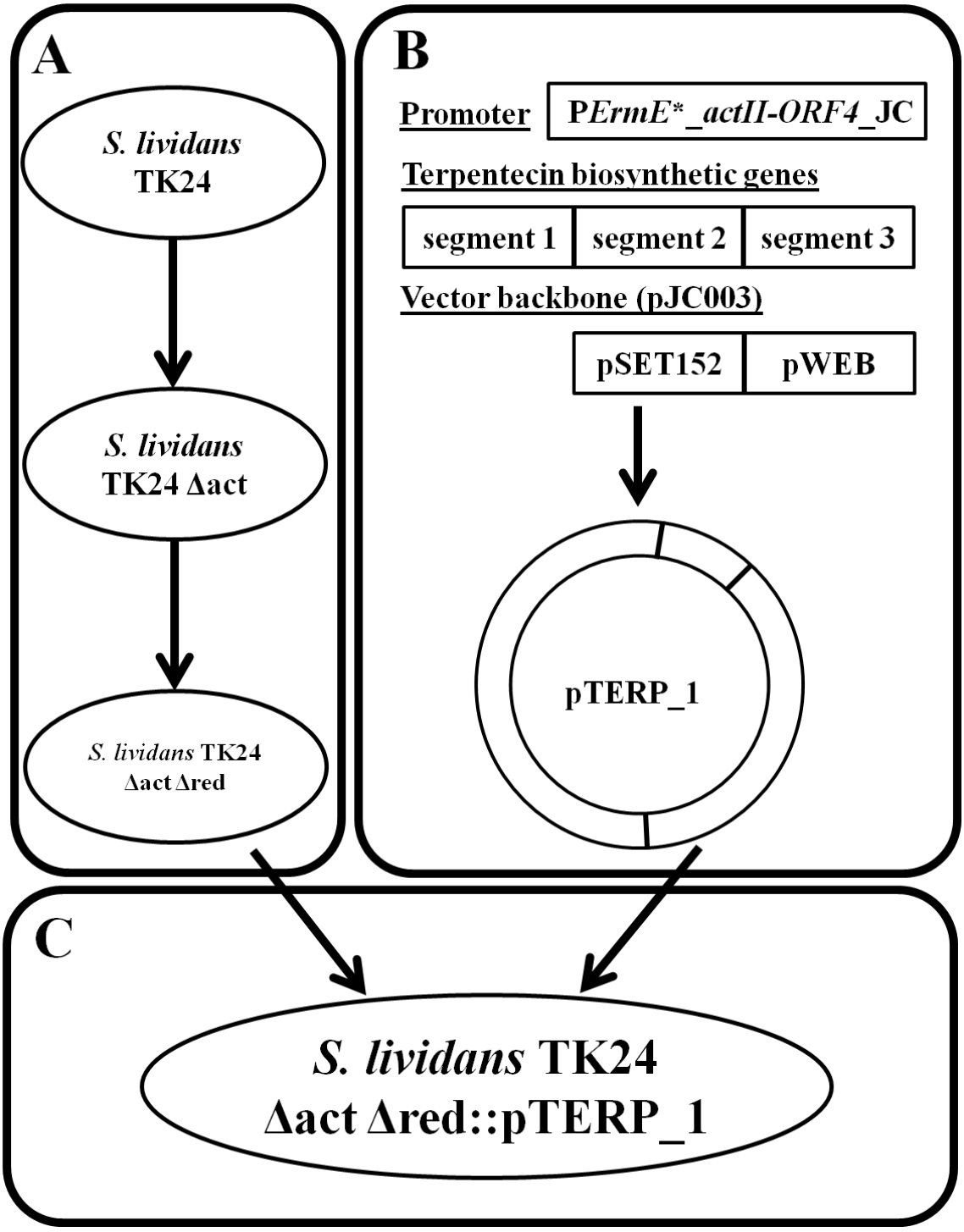


Figure 2.1. Schematic representation of the strategy to heterologously express terpentecin: host improvement strategy by knock-out of the actinorhodin (act) and undecylprodigiosin (red) biosynthetic gene clusters **(A)**; design of the integrating vector containing the biosynthetic genes and promoter **(B)**; introduction of the integrating vector into the genetically simplified host via conjugation **(C)**.

Table 2.1. Plasmids used and constructed throughout this study, Terp*:* terpentecin , Baf: bafilomycin, Apr: apramycin, Amp: ampicillin, Th: thiostrepton, Neo: neomycin.

Plasmid/strain	**Relevant characteristics/comments**	**Source/ reference**
pKC1139	Apr^R, rep^{UC}, rep^{ts}, *oriT, lacZα*, MCS, S*treptomyces* replicating cloning vector	[69]
pKC1139_act_KO	pKC1139 with act homologous arms, Actinorhodin deletion plasmid	This work
pKC1139_red_KO	pKC1139 with red homologous arms, Undecylprodigiosin deletion plasmid	This work
pSET152	Apr^R, rep^{UC}, *oriT*, $int^{\Phi C31}$, *attP, lacZα,* MCS, *Streptomyces* integrating cloning vector	[69]
pSEGMENT1	pSET152 with terp *ORF8-13*	This work
pSEGMENT2	pSET152 with terp *ORF13-1*	This work
pSEGMENT3	pSET152 with terp *ORF1-7*	This work
pWEB	Amp^R, $ori^{Co\ lE1}$, *cos*, ori^{SV40}, Neo^R	Epicentre®
pUC57	Amp^R, $rep^{(pMB1)}$, *lacZ*, MCS, cloning vector	Bio Basic
pJC003	Apr^R, *oriT*, $int^{\Phi C31}$, *attP*, amp^R, $ori^{Co\ lE1}$, *cos*	This work
pTERP_1	pJC003 with *terp* cluster and promoter	This work
pSET151	Th^R, Apr^R, *XylE+*, rep^{UC}, *oriT, lacZα*, MCS, *Streptomyces* non-replicating cloning vector	[69]
pBAFKO	pSET151 with homologous arms and *aa(3)IV*, bafilomycin suicide plasmid	This work
pTERPKO	pSET151 with homologous arms and *aa(3)IV*, terpentecin suicide plasmid	This work
E. coli 5α/10β	High-efficiency chemically competent host for cloning	New England Biolabs
E. coli ET12567:: pUZ8002	Donor strain during intergeneric conjugation	[70]
K. griseola MF730-N6	Native terpentecin producer	[67]
S. lividans TK24	*str-6*, $SLP2^-$, $SLP3^-$	John Innes Centre[71]
S. lividans TK24 ΔactΔred	Act^-, Red^-	This work
S. lividans TK24 ΔactΔred::pTERP1	*S. lividans* TK24 ΔactΔred with *terp* gene cluster under P*ErmE*_actII-ORF4*_JC	This work
S. lividans TK24 ΔactΔred::pJC003	*S. lividans* TK24 ΔactΔred with vector backbone only	This work
K. griseola Δbaf	$Terp^+$, Baf^-	This work
K. griseola Δterp	$Terp^-$, Baf^+	This work

Table 2.2. Sequences of primers used in this study, annealing location in sequence of pTERP_1, [a]annealing location refers to GenBank accession numbers AL939122, GG657756.1 and AP010968.1 for act, red and bafilomycin, respectively and according to the constructed plasmid pTerp_1 for terpentecin related primers, engineered restriction sites are underlined.

Primer	Sequence (5′→3′), restriction sites underlined	Anneals at[a]
M13F	TGTAAAACGACGGCCAGT	N/A
M13R	CAGGAAACAGCTATGACC	N/A
act_left_F	GGAAGCTTGCCGGAAATCCAGCGTATGC	158804
act_left_R	GCGTCCATGGCACCCCATCTCCCTTCGACC	159801
act_right_F	GTACCCATGGGATCAGCTCGCGGAGGATGT	165131
act_right_R	GCGTGGATCCGAACGTCCGCCTGGTCGAGA	166213
act_nested_F	GTCACCAGTGCGACTTCGGA	159679
act_nested _R	GACATCCTCCGCGAGCTGATC	165151
red_left_F	CTGGAAGCTTGCGACCGCGCACAGGACCTT	2045009
red_left_R	GCGTCTCGAGCTGGAGAACGTGCCGGCCGA	2046531
red_right_F	GCGTCTCGAGAAACGAAGCACCCCACCGGG	2078660
red_right_R	GTACGGATCCGGAGCGTCCCGTCAGAGCGT	2080181
TERP-SEG 1-F	GTCATATGGGATCCCGAGGAGGGAGCGA	9293
TERP-SEG 1-R	ATGTCGCAGCCGACTCCCAT	16062
TERP-SEG 2-F	ATGCTCCCGTCGTTCAGCCA	15901
TERP-SEG 2-R	GTACTAGTATCCAGGAATGTGCTAGGCG	18925
TERP-SEG 3-F	CTGGCCGACTATGTCGCTTT	18863
TERP-SEG 3-R	GTACTAGTGGCTGCTCCGGTCAGCACGC	25433
SEQTERP-1-d1-F	GGTGTTCGTGGCGGGAGTCG	9799
SEQTERP-1-d2-F	GCGATGATCTCGACCTTCTT	10400
SEQTERP-1-d3-F	GGGCGAGCACCGTCCGACCG	10995
SEQTERP-1-d4-F	CCAGTGCAGACGAGTGGCGA	11602
SEQTERP-1-d5-F	GACATCGCGTATCCGCTGCC	12203
SEQTERP-1-d6-F	ACCACCTCGCTTTCGCCACC	12801
SEQTERP-1-d7-F	ACTGAGCGAGGTGCTGCAGC	13397
SEQTERP-1-d8-F	GGAGCAAACACCCTCGGAGA	14006
SEQTERP-1-d9-F	CGGCCTACAGCTCCATCATT	14607
SEQTERP-1-d10-F	TACTACTCGTGGGGCCGGGAGAAG	15203
SEQTERP-1-d11-F	GTGAGTGACGCGGACAAGTT	15809
SEQTERP-1-d1-R	GAAGAGGTGGCAGTGCTCAT	15582
SEQTERP-1-d2-R	TGGCGCGCTTGATCTCGTGG	14965
SEQTERP-1-d3-R	ATATCGCCGCCGCACACTCCTC	14388
SEQTERP-1-d4-R	GCACCCAGAGCACCTCGAAG	13779
SEQTERP-1-d5-R	AACCCGACCGATCCCCACAG	13204
SEQTERP-1-d6-R	CTCCAGCTGATCAGGGTGAT	12577
SEQTERP-1-d7-R	TATGCCCTGTGTCAGGTTGG	11980
SEQTERP-1-d8-R	CACACTGATCGGCCTGGTCC	11376
SEQTERP-1-d9-R	CCACCTCCTGCATGGTCTGC	10778
SEQTERP-1-d10-R	TCGCGCCCGGCAGGTCGAGG	10175

Table 2.2 Sequences of primers used in this study (continued)

SEQTERP-1-d11-R	ACGTGCGTGGCAAGTCGGCA	9575
SEQTERP-2-d1-F	TGCTACGAACTGGCCCGGCA	16433
SEQTERP-2-d2-F	TGCCTGATCTCTGCCGCCGA	17037
SEQTERP-2-d3-F	AGGCTCGCCGTCCGTTACGA	17635
SEQTERP-2-d4-F	CGATTCTGGCCGGCGACGTT	18234
SEQTERP-2-d5-F	CTCACCGAGGAGCAGCGCCG	18830
SEQTERP-2-d1-R	TCGCAGGCGCAGCTCAGCAG	18456
SEQTERP-2-d2-R	ACAAGAAGTGCCCCTTTGCC	17852
SEQTERP-2-d3-R	TCAGGTCCCCACGTCTTGCG	17258
SEQTERP-2-d4-R	TCCTTCGGAATGGCGTAGCC	16656
SEQTERP-2-d5-R	TCGCAGCCGACTCCCATGAC	16059
SEQTERP-3-d1-F	ACGCACGAGCGATCATCCTC	19357
SEQTERP-3-d2-F	GAACAGCCGTCAACCCTGCT	19978
SEQTERP-3-d3-F	TGGTGGTCGCGGTGGTCAGC	20555
SEQTERP-3-d4-F	ACCCGGTGTCGTCGTCTCCT	21155
SEQTERP-3-d5-F	TCGTGGACGCCGCAGTCATC	21778
SEQTERP-3-d6-F	CTTTTTCAGGGACCCCTCCT	22387
SEQTERP-3-d7-F	ACAGATCTCCACCTGGCTGG	22987
SEQTERP-3-d8-F	ACATGGCGACCCTGGCTTCC	23591
SEQTERP-3-d9-F	GATGCGTGCGCATGTCCAAC	24192
SEQTERP-3-d10-F	CCGTCGACGTCATGGACTTC	24794
SEQTERP-3-d11-F	CCACAAGCGCCTCTACCGAG	25395
SEQTERP-3-d1-R	GCTGGTGGTAGCAGAACGCG	24961
SEQTERP-3-d2-R	GTGGTACTTGCCGATCTCGG	24360
SEQTERP-3-d3-R	TTCAGGACGTCGGCCACCAC	23783
SEQTERP-3-d4-R	AACGGATGTCGGCGGAGCTG	23165
SEQTERP-3-d5-R	GTCTCCTGCGCCGTGTTCAG	22560
SEQTERP-3-d6-R	CCCCGAGGGCTTGGCCGCGC	21962
SEQTERP-3-d7-R	AACTTCCGGCCGTCCTCGTG	21358
SEQTERP-3-d8-R	GCCGAGGGCGTTGCGCTCCG	20748
SEQTERP-3-d9-R	GCGAGCTGGACCCGGGTGGT	20156
SEQTERP-3-d10-R	AGCCGATGTCCAGCGCCTGT	19529
SEQTERP-3-d11-R	GAGACCGCATCGTCAGGAAT	18943
PROMOT-B-F	GACCGGTTCAGATTCCGCTGGTCGCACCCGTCA	9156
PROMOT-B-R	TGCGTCGACTCTAGAGATATCGCATGCCATATG ATCTCCCTTCGACCGCCGCT	9287
PROMOT-B-F-Con	GATTCCGCTGGTCGCACCCGTCA	9156
PROMOT-B-R-Con	ATCTCCCTTCGACCGCCGCT	9287
PROMOT-1-R	ATAAGTTGAATCTCACCACC	8391
VEC-pSET-F	P-CCGACTATTTGCAACAGTGC	6415

Table 2.2 Sequences of primers used in this study (continued)

VEC-pSET-R	P-CTTCCCGGGTGTCTCGCTAC	2361
TERP_ORF9_F	CGGGCGTACTCCTCGATCAT	11076
TERP_GGPPS_F	CCGATCGTCGGGTACCACTT	17978
TERP_GGPPS_R	AAGCGACATAGTCGGCCAGC	18881
TERP_HMG_coA_Red_F	CATCTCGGGAAACGTCGCCA	23244
TERP_HMG_coA_Red_R	CATGCGCACGCATCAGTTCG	24205
HrdB_F	CTCGAGGAAGAGGGTGTGAC	N/A
HrdB_R	TGCCGATCTGCTTGAGGTAG	N/A
bafKO_left_F	GCTTCGGAATTCGGGCGACCAGCTCCCGTTTG	8335543
bafKO_left_R	GTAGCATGCCCGCCCGGCGGCTGGACGTG	8336661
bafKO_right_F	GTAGCATGCTGGCTGGACACCGACCTCGA	8407725
bafKO_right_R	GACTGCAAGCTTGACGAGCTGGCGATCTGGAG	8408740
terpKO_left_F	GCTTCGGAATTCAGCAAGAGGAATTGGAGCCG	12000
terpKO_left_R	GTAGCATGCACAGGGCCGCTATCCGATCG	13025
terpKO_right_F	GTAGCATGCCCGAATCGGCCTACAGCTCC	14601
terpKO_right_R	GACTGCAAGCTTGTGACGGAGCCGGTGACGCT	15584
pJV177_aprR_F	CTAGCATGCAGCGTACGGCCCACAGAATG	N/A
pJV177_aprR_R	TTCGGCATGCTCATGAGCTCAGCCAATCGAC	N/A
TERP_602_F	CGACCTGATGCCGGGTGCGG	12595
TERP_1660_R	GCCGAGGAGAAGTACCTGCC	N/A

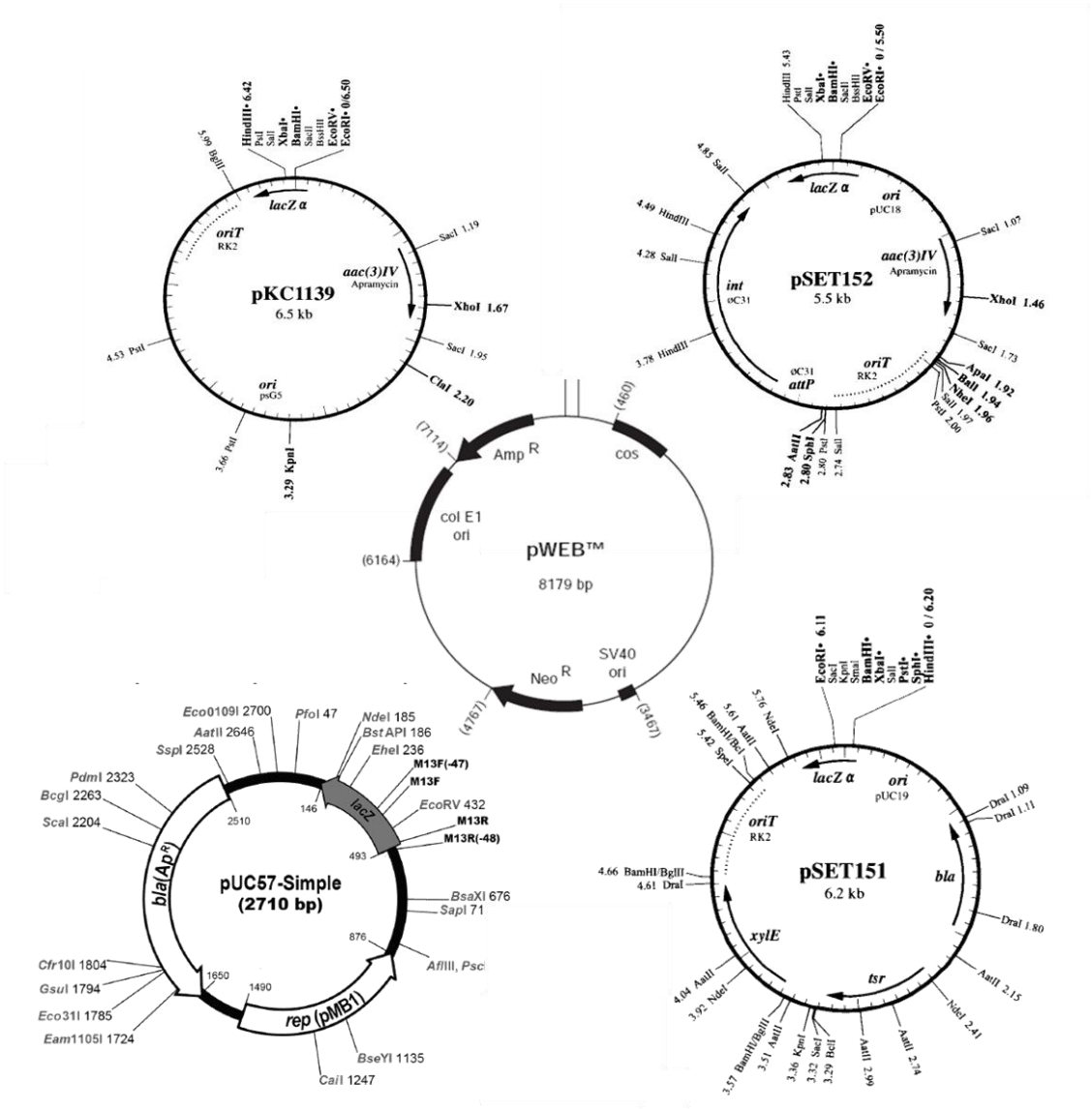


Figure 2.2. Maps of plasmids used in this study: pKC1139, pSET152, pWEB, pUC57 and pSET15; reproduced with permission from Keiser et al.[71], Epicentre® pWEB user manual and www.genscript.com/gsfiles/techfiles/GenScript_pUC57-Simple_Plasmid_Map.pdf.

Table 2.3. Recipes for selected media used in this study.

Medium	**Components (per L of diH_2O, unless specified otherwise)**
Seed medium	10 g glucose 10 g yeast extract
Maltose-yeast-malt (MYM)	4 g maltose 4 g yeast 10 g malt extract
Mannitol-soy	20 g mannitol 20 g soy flour tap water
Lean medium	0.4 g glucose 0.8 g galactose 0.8 g maltose 1.6 g dextrin 0.8 g Bacto™-Soytone 0.3 g $(NH_4)_2SO_4$
Yeast extract-malt extract (YEME) + 34% sucrose + 0.5% glycine	3 g yeast extract 3 g Bacto™-Peptone 3 g malt extract 10 g glucose 340 g sucrose 5 mM (final) $MgCl_2 \cdot 6H_2O$ 0.5% (final) glycine
International *Streptomyces* project medium 3 (ISP3)	20 g oatmeal (boiled in 1 L diH_2O for 20 min and filtered) 1 mL trace element solution[a]
Tryptic soy broth (TSB)	17 g pancreatic digest of casein (tryptone) 3 g Bacto™-Soytone 5 g NaCl 2.5 g K_2HPO_4 2.5 g dextrose
[a]Trace element solution	1 g $FeSO_4 \cdot 7H_2O$ 1 g $MnCl_2 \cdot 4H_2O$ 1 g $ZnSO_4 \cdot 7H_2O$

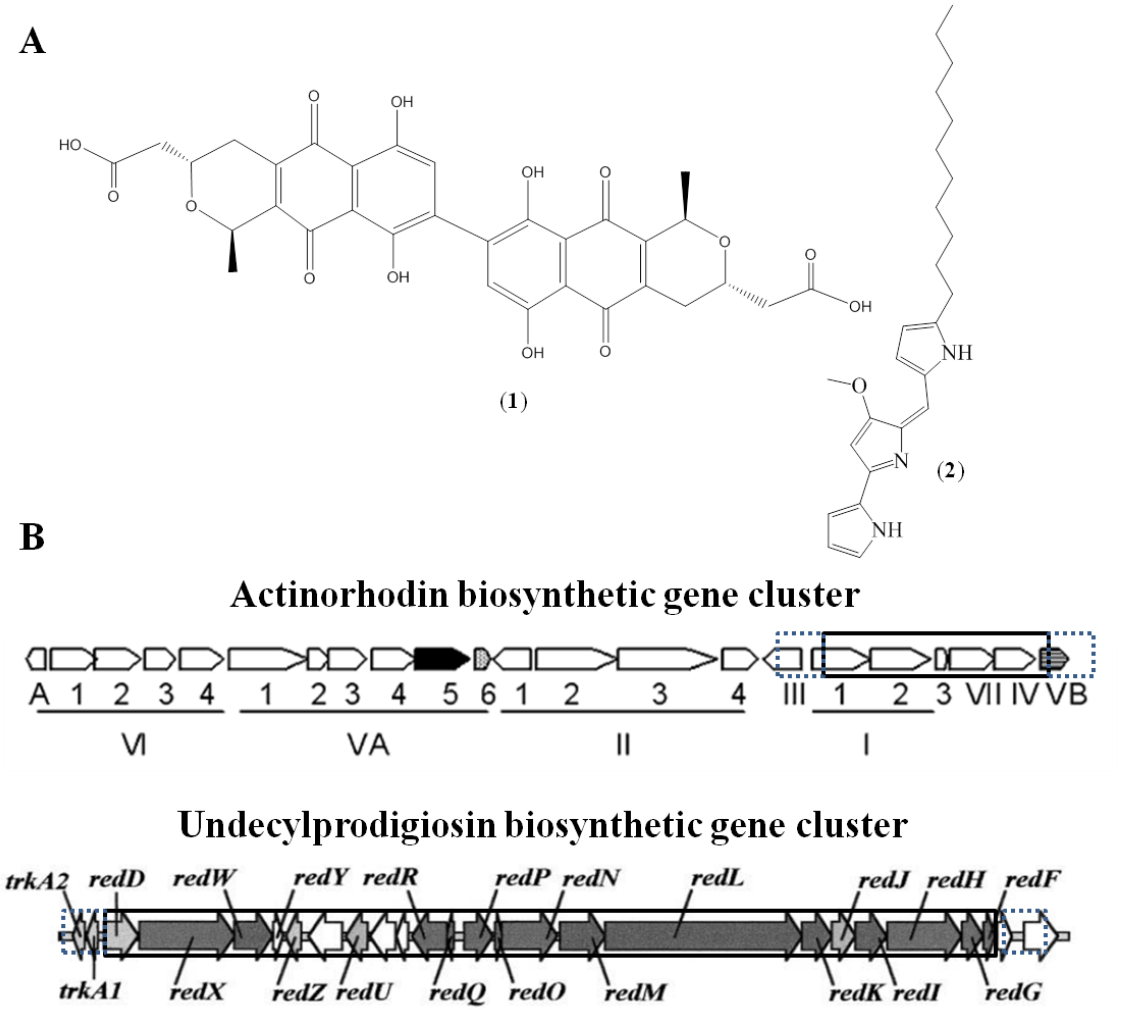


Figure 2.3. Structures of actinorhodin (**1**) and undecylprodigiosin (**2**), two endogenous compounds produced by *S. lividans* TK24 **(A)** and arrangement of the biosynthetic gene cluster of the actinorhodin and undecylprodigiosin in the closely related bacterium, *Streptomyces coelicolor* A3(2). Homologous arms and regions removed are shown in dotted and black boxes, respectively, modified from Okamoto et al.[72] and Cerdeno et al.[73] **(B).**

clusters were deleted from the genome using a double-crossover gene replacement strategy[71].

Act deletion plasmid construction

To delete a 5.33 kilobase pair (kb) region of the *act* cluster containing *act I ORF1-3*, *actVII* and *actIV* (Figure 2.3 (B)), a 997 base pair (bp) fragment upstream of *actI-ORF1* from bases 158,804 to 159,801 and a 1082 bp fragment downstream of *actVB* from bases 165,131 to 166,213 were amplified by polymerase chain reaction (PCR) using *S. lividans* TK24 genomic DNA (gDNA) as template, KOD Hot Start DNA Polymerase and primer pairs act_left_F/act_left_R and act_right_F/act_right_R, respectively according to the conditions listed in Table 2.4. The act_left and act_right amplicons were restricted with HindIII/NcoI and NcoI/BamHI, respectively and purified from agarose gels following electrophoresis. The arms were cloned into the *E. coli*/*Streptomyces* shuttle vector, pKC1139 (digested with HindIII/BamHI), in a three-way ligation. Ligation was performed using T4 DNA ligase according to the manufacturer's instructions and a 3:1 insert to vector ratio, giving the plasmid pKC1139_act_KO.

A portion of the ligation was used to transform *E. coli* 5α and recombinant colonies were selected using Luria-Bertani (LB) agar containing 50 µg/ml apramycin according to standard protocols[68]. Transformants were screened for the presence of the desired plasmid by colony PCR. Briefly, a small portion of *E. coli* colonies were suspended in 25 µL of dimethyl sulfoxide (DMSO) using a sterile tooth pick and the resulting suspension was used as template DNA in PCR reactions using EconoTaq® conducted using the primers act_left_F and act_right_R (Table 2.2). Amplification of a 2500 bp band indicated colonies contained the desired plasmid. Colonies testing positive in the colony PCR assay were cultured overnight in LB broth containing apramycin and plasmid DNA (pDNA) was isolated and subjected to restriction analysis using XhoI, BamHI and BamHI/NcoI. Plasmids showing the expected banding pattern for all digests were used to transform *E. coli* ET12567(pUZ8002) cells. Transformants were plated on LB agar containing chloramphenicol (25 µg/mL), kanamycin (50 µg/mL) and apramycin (50 µg/ml).

Table 2.4. Polymerase chain reaction conditions used throughout this study, times are reported in min:s.

Amplicon name	Initial denaturation	Amplification				Final elongation
		Cycles	Denaturation	Annealing	Polymerization	
act_left	95˚C for 02:00	25	95˚C for 00:20	60-70˚C for 00:30	70˚C for 00:30	N/A
act_right	95˚C for 02:00	25	95˚C for 00:20	60-70˚C for 00:30	70˚C for 00:30	N/A
red_left	95˚C for 02:00	35	95˚C for 00:20	74.2˚C for 00:10	70˚C for 00:30	70˚C for 05:00
red_right	95˚C for 02:00	35	95˚C for 00:20	75.6˚C for 00:10	70˚C for 00:30	70˚C for 05:00
screen_act	95˚C for 02:00	35	95˚C for 00:20	60-65˚C for 00:30	70˚C for 00:45	N/A
screen_red	95˚C for 02:00	30	95˚C for 00:20	74.2˚C for 00:10	70˚C for 01:00	N/A
confirm_act_KO	95˚C for 02:00	30	95˚C for 00:20	59.6˚C for 00:10	70˚C for 02:30	N/A
confirm_red_KO	95˚C for 02:00	30	95˚C for 00:20	74.2˚C for 00:10	70˚C for 01:00	N/A
segment 1	95˚C for 02:00	30	95˚C for 00:20	62˚C for 00:10	70˚C for 02:11	N/A
segment 2	95˚C for 02:00	30	95˚C for 00:20	62˚C for 00:10	70˚C for 01:00	N/A
segment 3	95˚C for 02:00	30	95˚C for 00:20	60.8˚C for 00:10	70˚C for 02:03	70°C for 05:00
promoter-frag_B	95˚C for 02:00	30	95˚C for 00:20	75.8˚C for 00:10	70˚C for 0:10	N/A

Table 2.3 (continued). Polymerase chain reaction conditions used throughout this study, times are reported in min:s.

vec-pSET	95˚C for 02:00	25	95˚C for 00:20	60.2˚C for 00:10	70˚C for 01:40	N/A
confirm_seg_2-3	98˚C for 2:00	25	98˚C for 00:20	62.3˚C for 00:15	72˚C for 4:00	72˚C for 05:00
confirm_pro-seg_1	98˚C for 2:00	25	98˚C for 00:20	69.8˚C for 00:15	72˚C for 1:00	72˚C for 05:00
RT_ORF9	94˚C for 2:00	30	94˚C for 00:15	60.5˚C for 00:15	72˚C for 0:30	72˚C for 5:00
RT_cyclase	94˚C for 2:00	30	94˚C for 00:15	62.9˚C for 00:15	72˚C for 1:00	72˚C for 5:00
RT_GGPPS	94˚C for 2:00	30	94˚C for 00:15	64.5˚C for 00:15	72˚C for 1:00	72˚C for 5:00
RT_HMG-CoA_R	94˚C for 2:00	30	94˚C for 00:15	64.5˚C for 00:15	72˚C for 1:00	72˚C for 5:00
RT_hrdB	94˚C for 2:00	30	94˚C for 00:15	62.4˚C for 00:15	72˚C for 0:30	72˚C for 5:00
bafKO_left	94˚C for 2:00	30	94˚C for 00:15	71.7˚C for 00:15	72˚C for 1:20	72˚C for 5:00
bafKO_right	94˚C for 2:00	30	94˚C for 00:15	73.1˚C for 00:15	72˚C for 1:00	72˚C for 5:00
terpKO_left	94˚C for 2:00	30	94˚C for 00:15	71.1˚C for 00:15	72˚C for 1:00	72˚C for 5:00
terpKO_right	94˚C for 2:00	30	94˚C for 00:15	71.1˚C for 00:15	72˚C for 1:00	72˚C for 5:00
aprR	94˚C for 2:00	30	94˚C for 00:15	71.3˚C for 00:15	72˚C for 1:00	72˚C for 5:00
screen_bafKO	94˚C for 2:00	30	94˚C for 00:15	73.6˚C for 0:15	72˚C for 3:00	72˚C for 5:00

Table 2.3 (continued). Polymerase chain reaction conditions used throughout this study, times are reported in min:s.

conf_bafKO_1	94˚C for 2:00	30	94˚C for 00:15	73.6˚C for 0:15	72˚C for 3:00	72˚C for 5:00
conf_bafKO_2	94˚C for 2:00	30	94˚C for 00:15	73.1˚C for 00:15	72˚C for 1:20	72˚C for 5:00
conf_terpKO	94˚C for 2:00	30	94˚C for 00:15	61.1˚C for 00:15	72˚C for 4:00	72˚C for 5:00

Deletion of actinorhodin genes in *S. lividans* TK24

The deletion plasmid, pKC1139_act_KO, was delivered into *S. lividans* TK24 via intergeneric conjugation between *S. lividans* TK24 spores and *E. coli* ET12567/pUZ8002 harbouring pKC1139_act_KO according to published protocols[71]. Mixtures of *E. coli::*ET12567/pUZ8002/pKC1139_act_KO and *S. lividans* TK24 spores were plated on mannitol-soy agar containing 10 mM $MgCl_2$. Following overnight incubation at 30°C, *S. lividans* TK24 exconjugants containing pKC1139_act_KO were selected for by overlaying plates with 1 mL of water containing 0.5 mg nalidixic acid (to inhibit the growth of *E. coli*) and 1 mg apramycin and plates were incubated for an additional 16 h. Apramycin resistant exconjugants were subcultured twice on mannitol-soy and maltose-yeast extract-malt extract (MYM) containing apramycin and nalidixic acid and incubated at 30°C. Due to the temperature sensitive replication function of the plasmid, colonies growing at temperatures above 34˚C will only be resistant to apramycin if a single crossover event has occurred resulting in genomic integration of the plasmid; organisms were thus cultured at 37˚C on mannitol-soy containing apramycin. To select for a second crossover and removal of the plasmid from the host, colonies from the aforementioned plate were inoculated in 10 mL MYM broth containing no antibiotic and incubated at 30°C and 200 revolutions per minute (rpm). One hundred microliters of each culture was plated on mannitol-soy containing no antibiotic and grown at 30°C until colonies sporulated. It is expected that 50% of the resulting spores would undergo gene deletion via a double crossover, while the other half would revert to the original genotype. Spores were harvested and plated for single colonies on mannitol-soy. The sensitivity of individual colonies to apramycin was evaluated by sequentially plating each colony on mannitol-soy with and without apramycin. Apramycin sensitive colonies were kept and their genotype and phenotype were confirmed as described below, giving rise to *S. lividans* TK24 Δact.

The gene deletion procedure was repeated, using *S. lividans* TK24 Δact, where a deletion plasmid was constructed from two 1.5 kb PCR amplified homologous regions at the extremities

of the 35 kb undecylprodigiosin biosynthetic gene cluster (Figure 2.3). The homologous arms were cloned into the HindIII and BamHI restriction sites of pKC1139 and a XhoI restriction site was engineered in order to ligate the left and right homologous arms. The steps were as described above for the Δact strain, using primers red_left_F and red_left_R to amplify the left homologous arm (2045009-2046531 = 1522 bp), red_right_F and red_right_R to amplify the right homologous arm (2078660-2080181 = 1521 bp) and finally using red_left_F and red_right_R primers in order to screen via PCR for *E. coli* transformants containing the deletion plasmid. EcoRV, XhoI, KpnI and PstI were used to confirm correct assembly of pKC1139_red_KO. Conjugation with *S. lividans* TK24 Δact was performed using *E. coli* ET12567 (pUZ8002) as the plasmid donor strain and double crossover apramycin-sensitive exconjugants were kept, giving rise to the genetically simplified strain *S. lividans* TK24 ΔactΔred.

Confirming the genotype and phenotype of knock-out strains

To confirm the loss of the 5.33 kb region of the act cluster, two primers were designed (Table 2.2): act_nested_F which annealed 25 bp upstream of the end of the act_left arm, and act_nested_R which annealed 20 bp downstream of the start of the act_right arm. These primers amplified a 5472 bp fragment in wild-type colonies and a 145 bp amplicon in mutant colonies. Amplicons using the above primers were obtained using the PCR conditions listed in Table 2.4.

Similarly, red_left_F and red_right_R primers (Table 2.2) were used to amplify the confirm_red_KO amplicon according to the conditions in Table 2.4 in order to confirm the loss of the 32.13 kb portion of the undecylprodigiosin biosynthetic gene cluster; an amplicon from revertant colonies could not be observed due to the large distance (35 kb) between the primers' annealing location but a 3042 bp amplicon resulted from colonies having undergone gene deletion.

In addition to the genetic characterization of the knock-out strains, the phenotype was also investigated. Firstly, pigmentation was visually assessed following a 14-day fermentation in

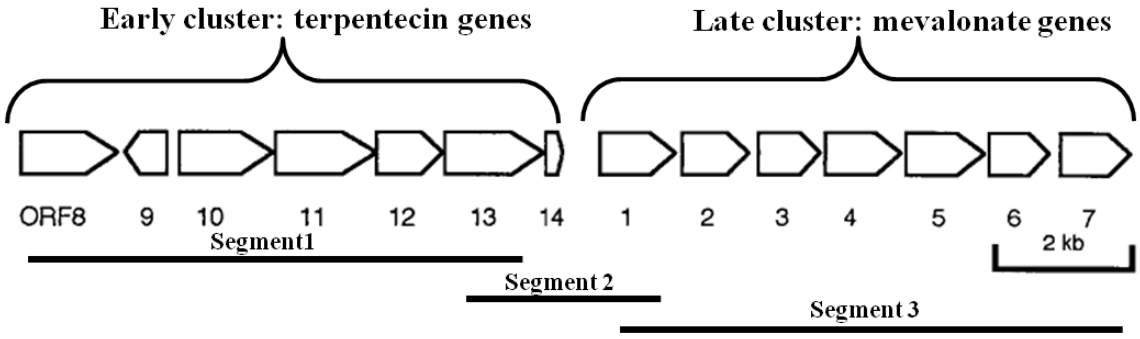
MYM broth. Secondly, taking advantage of the colour changes of act and red in response to different pH, 10 drops of 5 M hydrochloric acid (HCl) and 10 drops of 5 M sodium hydroxide (NaOH) were added to 3 mL of cultures of *S. lividans* TK24 and *S. lividans* TK24 Δact strain 6.1. Thirdly, quantification of each compound was achieved spectrophotometry by modifying the methods of Borodina et al.[74] Actinorhodin was measured from six Δact strains, the revertant strain 18.24 and wild-type strain as a positive control. Similarly, red was quantified from six ΔactΔred strains using Δact 6.1 and TK24 as positive controls. Organisms were grown in MYM at 30°C and 200 rpm for 14 d. Briefly, 5 mL of culture was treated with 10 mL of 2 M NaOH and 10 mL acidified methanol (MeOH) (to pH 1.5 with HCl) for act and red, respectively and shaken at 150 rpm overnight. The samples were centrifuged at 4000 X *g* for 10 min, 1 mL of the supernatant removed, centrifuged again at 10000 X *g* for 2 min and the absorbance (*A*) of the supernatants measured using a NanoDrop spectrophotometer (Fisher Scientific, Ottawa, ON) at 640 nm and 530 nm for act and red, respectively. Concentrations of act and red were calculated using the rearranged Beer-Lambert equation, $c=A/(\varepsilon \cdot l)$, using 0.1 cm for path length (l) and 25320 $M^{-1}cm^{-1}$ and 100500 $M^{-1}cm^{-1}$ as extinction coefficients (ε) for act and red, respectively. MYM medium was treated the same and used to blank the spectrophotometer.

An analysis of variance (ANOVA) was performed to determine if there was a significant difference in the concentration of each compound between groups (different strains used); an α level of significance of 0.05 and sample sizes of 5 and 3 were used for act and red, respectively. Residuals were tested for normality and homoscedasticity.

2.2.2 Construction of pTERP_1: The terpentecin biosynthetic gene cluster construct

Amplification of the terpentecin biosynthetic genes in three individual segments

The terpentecin biosynthetic gene cluster published by Dairi et al. in 2001[64] (Figure 2.4) was amplified in three segments via PCR using KOD Hot Start DNA Polymerase according to the manufacturer's instructions using *K. griseola* MF730-N6 genomic DNA as a template and the conditions listed below. Restriction sites were incorporated into the primers of each segment



Open reading frame	Predicted protein from gene annotation	Location[a]
ORF 1	Geranylgeranyl diphosphate synthase	17849-18895
ORF 2	Mevalonate kinase	18976-19979
ORF 3	Probable mevalonate diphosphate decarboxylase	20112-21014
ORF 4	Probable phosphomevalonate kinase	21012-22100
ORF 5	Type II IPP isomerase	22097-23191
ORF 6	HMG-CoA reductase	23188-24249
ORF 7	Probable HMG-CoA synthase	24380-25422
ORF 8	Probable transmembrane efflux protein	9525-10963
ORF 9	Unknown protein	11018-11530*
ORF 10	Putative P450	11771-13021
ORF 11	Diterpene cyclase	13044-14543
ORF 12	Diterpene cyclase	14543-15478
ORF 13	Putative P450	15542-16990
ORF 14	Probable ferrodoxin	16998-17195

Figure 2.4. Graphical representation and annotation of the terpentecin biosynthetic gene cluster. Modified from Dairi et al.[64]; [a]location of genes in the sequence of the plasmid pTERP_1, *complementary strand translation.

in order to facilitate downstream steps and primers were ordered with a 5' phosphate in order to perform blunt-end ligations of the amplicons with dephosphorylated linearized vector DNA.

The 5′ segment (segment 1), which contained the first 6770 bp of the cluster, was amplified using TERP-SEG 1-F and TERP-SEG 1-R primers (Table 2.2), segment 2, which contained base pairs 6650 to 9633 of the cluster was amplified using TERP-SEG 2-F and TERP-SEG 2-R primers (Table 2.2) and segment 3, which contained base pairs 9575 to 16133 of the cluster was obtained using TERP-SEG 3-F and TERP-SEG 3-R primers (Table 2.2) according to the conditions listed in Table 2.4.

Each amplicon was gel-purified using the Zymoclean gel DNA recovery kit (Zymo Research) and was cloned into pSET152[69] that had previously been restricted with EcoRV and dephosphorylated with Antarctic Phosphatase in order to avoid vector recircularization. Blunt ligations were performed using T4 DNA ligase and chemically competent *E. coli* 10β was transformed with 5 µL of the ligation mixtures and plated on LB medium containing 50 µg/mL apramycin.

Plasmid DNA was extracted from five colonies from each transformation using the QIAprep Spin Miniprep Kit (Qiagen, Toronto, ON) and pDNA was subjected to restriction analysis using XbaI/NheI and NcoI to confirm proper assembly and direction of each amplicon in pSET152. Plasmid DNA from a clone showing the expected restriction pattern was sent for sequencing to ensure the cloned biosynthetic gene cluster segments contained no errors. This was done using the M13F and M13R primers as well as primers designed to anneal at ~600 bp intervals in order to sequence the entire length of the segments; primer sequences are listed as SEQTERP in Table 2.2. The resulting sequences were assembled and compared to the published sequences (accession numbers AB037907.1 and AB048795.1)[64, 75] using the software *Contig*Express® (Life Technologies, Burlington, ON). Any DNA regions containing discrepancies were resequenced using fresh pDNA preparations and amplicons from genomic DNA using EconoTaq® and the following conditions: 95°C for 2 min, 30 cycles of 95°C for 20

s, X°C for 10 s, 70°C for Y s and a final extension of 70°C for 5 min, where the variables are listed Table 2.5 along with the various primers used to cover all inconsistencies; giving rise to pSEGMENT1, pSEGMENT2 and pSEGMENT3.

To obtain the segments required for the assembly of the desired terpentecin biosynthetic gene cluster construct, segment one was obtained by restricting pSEGMENT1 with NdeI and NsiI, segment two by restricting pSEGMENT2 with AclI and NsiI, and segment three by restricting pSEGMENT3 with AclI and EcoRI. Restricted fragments were subsequently subjected to gel electrophoresis and bands of sizes 6.7 kb, 3 kb and 6.5 kb, respectively, were purified from the agarose.

Construction of the constitutive promoter *ErmE* actII-ORF4* JC.

The previously described constitutive promoter *ErmE** in conjunction with *actII-ORF4* (Figure 2.5) was used to ensure transcription of the heterologous genes[76]. Multiple restriction sites were incorporated at both ends of the promoter in order to increase its versatility and since possible GTG start codons were present within the promoter sequence (Figure 2.6), an engineered stop codon was included to cease transcription of the complementary strand, rendering transcription of the heterologous genes more efficient. The complete sequence of the promoter constructed in this study is illustrated in Figure 2.6. The promoter was constructed from two fragments (A and B).

Fragment A was synthesized by Bio Basic Inc. (Markham, ON) and contained the first 1060 bp of the promoter. It was provided in the SmaI site of pUC57. In order to clone fragment B into pUC57 containing fragment A, plasmid DNA was digested with AgeI and SalI. The 2.7 kb fragment was then gel-purified and dephosphorylated using Antarctic Phosphatase. Fragment B was obtained by PCR using *Streptomyces lividans* TK24 gDNA as a template and KOD Hotstart DNA Polymerase to amplify the 130 bp *actIII* to *actI* intergenic region with an AgeI restriction site engineered in the PROMOT-B-F primer (Table 2.2) according to the conditions in Table 2.4. The product was column purified, restricted with AgeI and SalI and column purified.

Table 2.5. Components and conditions used to generate 10 amplicons from gDNA of *K. griseola* MF730-N6 in order to resequence all 13 regions of discrepancy between the cloned segments and the published terpentecin gene cluster sequences.

Amplicon	Primer pair	Annealing temperature X (°C)	Polymerization Y (s)
gDNA seg1-1	SEQTERP-1-d5-F SEQTERP-1-d5-R	58.5	12
gDNA seg1-2	SEQTERP-1-d6-F SEQTERP-1-d3-R	62.9	30
gDNA seg2-1	SEQTERP-2-d2-F SEQTERP-2-d2-R	58.5	12
gDNA seg3-1	SEQTERP-3-d1-F SEQTERP-3-d8-R	59.7	27
gDNA seg3-2	SEQTERP-3-d2-F SEQTERP-3-d7-R	59.7	27
gDNA seg3-3	SEQTERP-3-d4-F SEQTERP-3-d5-R	59.7	27
gDNA seg3-4	SEQTERP-3-d5-F SEQTERP-3-d5-R	60.9	11
gDNA seg3-5	SEQTERP-3-d8-F SEQTERP-3-d2-R	57.8	11
gDNA seg3-6	SEQTERP-3-d9-F SEQTERP-3-d1-R	57.8	11
gDNA seg3-7	SEQTERP-3-d10-F TERP-SEG 3-R	57.8	11

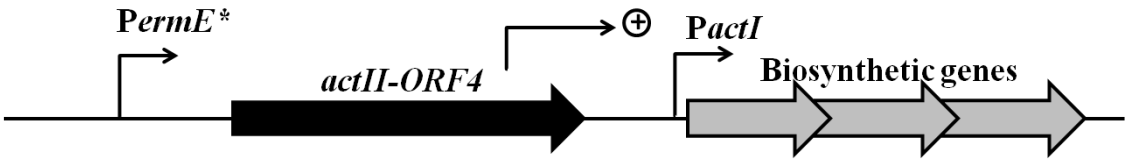


Figure 2.5. Design of the efficient promoter used to drive the transcription of the heterologous biosynthetic genes; adapted from Wilkinson et al.[76]

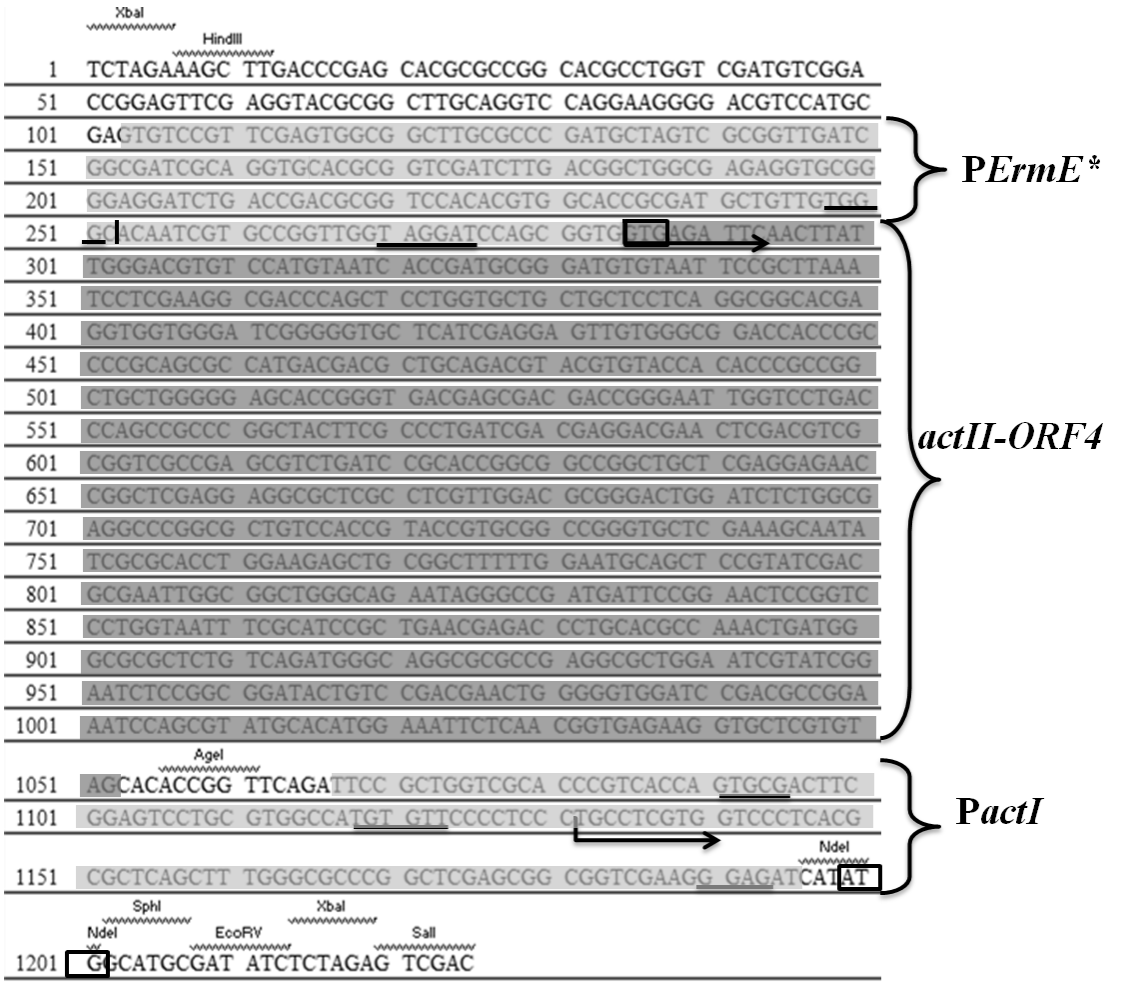


Figure 2.6. Sequence of P*ErmE*_actII-ORF4*_JC; start codons are boxed, the vertical line represents the 3 nucleotide deletion to give rise to P*ermE**, underligned regions are -10 and -35 ribosomal binding sites and arrows indicate promoter initiation points.

Fragments A and B were ligated using T4 DNA ligase (4:1 insert:vector ratio) and used to transform *E. coli* 10β competent cells, which were plated on LB plates containing 100 µg/mL ampicillin. Five transformants were screened by restricting pDNA with EcoRI, EcoRV, NdeI and BamHI. A plasmid exhibiting the expected banding pattern was named pUC57_P*ErmE*_actII- ORF4*_JC and completely sequenced to confirm no mutations had been introduced into the sequence. The promoter (1.2 kb) was removed from the vector by restriction with NdeI and EcoRI and gel-purified as described previously.

Construction of a new *Streptomyces* integrating cosmid vector: pJC003

An integrating plasmid-cosmid hybrid was designed in order to accommodate our large DNA fragment. This was composed of a 4 kb amplicon from pSET152 to allow antibiotic selection in *Streptomyces* (*aac(3)IV;* apramycin resistance), transfer of the plasmid during intergeneric conjugation (*oriT*, origin of transfer) and integration into the genome of *Streptomyces (ϕC31 integrase* and *attP*) and a 4 kb fragment from pWEB (Epicentre® via Mandel Scientific Company Inc., Guelph, ON) to provide stability of a large fragment and antibiotic selection (*bla*; ampicillin resistance) in *E. coli* (*colE1 ori*)*;* the portions used from each plasmid are highlighted in Figure 2.7 (A). Region 1657-5711 of pSET152 (according to GenBank AJ414670.1) was amplified using primers VEC-pSET-F/VEC-pSET-R (Table 2.2) using KOD Hot Start DNA Polymerase and the conditions listed in Table 2.4 and amplicons were column purified.

Linear pWEB from the Epicentre® Cosmid cloning kit was phosphorylated using T4 polynucleotide kinase (NEB) and recircularized using T4 DNA ligase. Circularized pWEB was used to transform and *E. coli* 10β and pDNA was purified and digested with HpaI, leaving blunt ends. The HpaI fragment was column purified and dephosphorylated using Antarctic Phosphatase. Gel-purification of the 4 kb band was not necessary because only plasmids formed from the correct fragment would result in resistance to ampicillin and apramycin.

The 4 kb fragment from pSET152 and 4 kb fragment from pWEB were ligated using a 4:1

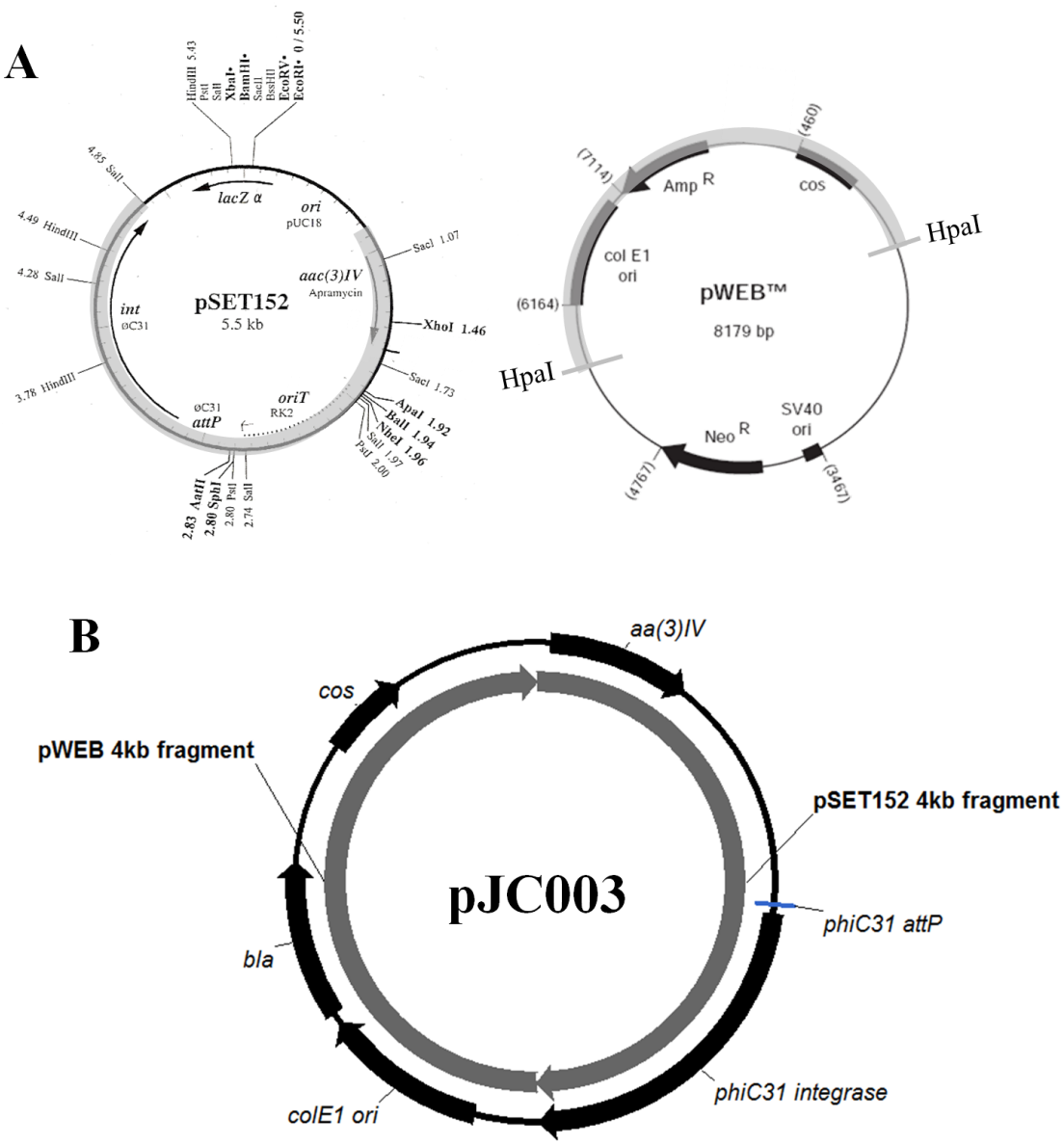


Figure 2.7. Vector maps for pSET152 and pWEB used to construct the new *Streptomyces* integrating vector pJC003; the portion amplified by PCR (pSET152) and obtained through HpaI digest (pWEB) are highlighted in grey **(A)** and vector map for pJC003 with originating fragments and genes in grey and black, respectively **(B)**.

insert to vector (pSET:pWEB) ratio using T4 DNA ligase and *E. coli* 10β was transformed with 5 µL of the ligation reaction. Transformants were selected on LB agar containing 50 µg/mL apramycin and 100 µg/mL ampicillin. Plasmid DNA from transformants was screened with EcoRV, NheI, HindIII and AclI (individually). The correctly constructed plasmid was named pJC003 (Figure 2.7 (B)).

To generate the vector for the heterologous expression of terpentecin genes, pJC003 was restricted with EcoRI-HF. Following gel-purification of the 8 kb band, the DNA was dephosphorylated and the Antarctic Phosphatase enzyme was inactivated by incubating the reaction at 70˚C for 5 min.

Assembly of the entire terpentecin biosynthetic gene cluster in pJC003, resulting in *E. coli*::pTERP_1

The three segments of the terpentecin gene cluster and P*ErmE**_actII-ORF4_*JC were cloned into pJC003 using the Invitrogen quick ligation kit (Life Technologies). The ligation was performed using a 6:1 insert to vector ratio using the quantities of DNA shown in Table 2.6. DNA used for the ligation reactions was concentrated from 45 µL to 15 µL using a Millipore centrifugal filter (nominal molecular weight limit: 30,000, Fisher Scientific) following the manufacturer's protocol. Five microliters of the ligation reaction was then used to transform chemically competent *E. coli* 10β, which was plated on LB containing 50 µg/mL apramycin and 100 µg/mL ampicillin.

Transformants were screened by colony PCR, as mentioned previously, in order to identify colonies containing all 5 components. Initially, 171 colonies were screened using SEQTERP-2-d3-F and SEQTERP-3-d7-R primers (Table 2.2), amplifying DNA at the junction of segments 2 and 3 (amplicon confirm_seg_2-3) using the conditions listed in Table 2.4. Colonies containing the expected 3.9 kb band were then screened by amplifying DNA at the junction of segments 1 and 2 (confirm_seg_1-2 amplicon) using the conditions listed in Table 2.4 and primers SEQTERP-1-d10-F and SEQTERP-2-d5-R (Table 2.2). Colonies showing the expected 856 bp band were grown overnight at 37˚C and 200 rpm in 10 mL LB containing

Table 2.6. Steps involved in the formation of each piece required for the formation of the terpentecin biosynthetic gene cluster construct.

	Segment 1	**Segment 2**	**Segment 3**	**Promoter**	**Vector**
Step 1	Extract plasmid DNA	Extract plasmid DNA	Extract plasmid DNA	Extract plasmid DNA	Extract plasmid DNA
Step 2	Cut NdeI + NsiI + XhoI	Cut AclI	Cut AclI + EcoRI-HF + XhoI	Cut EcoRI-HF + NdeI	Cut EcoRI-HF
Step 3	Gel-purify 6.77 kb band	Column purify	Gel-purify 6.56 kb band	Gel-purify 1.2 kb band	Gel-purify 8.1 kb band
Step 4		Cut NsiI			Dephosphorylate
Step 5		Gel-purify 3 kb band			Heat inactivate
Amount of DNA used	208 ng	92 ng	202 ng	37 ng	50 ng

apramycin and pDNA was purified for further analysis.

Purified pDNA from plasmids was sequenced using primers SEQTERP-1-d11-F, SEQTERP-1-d11-R, SEQTERP-2-d5-F, SEQTERP-2-d5-R, SEQTERP-3-d11-F, SEQTERP-3-d11-R, PROMOT-1-R, PROMOT-B-F (Table 2.2) which annealed at locations near junctions between segments in order to further confirm the proper location and orientation of each segment. The same pDNA was also subjected to a restriction digest analysis using the following enzymes: NcoI, AclI, SacI, HindIII, BsrGI, NdeI/NheI, NsiI, NsiI/NdeI, and finally NotI/NdeI and fragments compared to *in silico* digests, resulting in pTERP_1.

2.2.3 Heterologous expression of terpentecin in *S. lividans* TK24 ΔactΔred

Introduction of pTERP_1 into *S. lividans* TK24 ΔactΔred

E. coli ET12567::pUZ8002 was transformed with pTERP_1 and plated on LB containing 50 µg/mL apramycin, 50 µg/mL kanamycin and 25 µg/mL chloramphenicol. pTERP_1 was transferred from *E. coli* ET12567::pUZ8002 to *S. lividans* TK24 ΔactΔred via intergeneric conjugation according to published protocols[71]. Briefly, a mixture of *E. coli* ET12567::pUZ8002::pTERP_1 and *S. lividans* TK24 ΔactΔred spores were plated on mannitol-soy agar containing 10 mM $MgCl_2$ and incubated overnight at 30°C. Plates were then overlaid with 1 mL of water containing 0.5 mg nalidixic acid and 1 mg apramycin and the plates incubated for another 16 h. To select for colonies that contained pTERP_1, potential exconjugants were subcultured twice on mannitol-soy containing apramycin and nalidixic acid at 30°C. Colonies were then cultured, preserved in 25% glycerol and stored at -80˚C. This process was repeated using pJC003 in order to provide a control where *S. lividans* TK24 ΔactΔred contained solely the vector portion of the construct and no promoter or terpentecin related genes to give rise to *S. lividans* TK24 ΔactΔred::pJC003.

Colony PCR was performed on apramycin resistant exconjugants using PROMOT-B-F-Con and SEQTERP-1-d10-R primers (Table 2.2) and the cycling conditions listed in Table 2.4 to amplify the confirm_pro-seg1 and confirm_seg_2-3 amplicons described above, ensuring that

all three terpentecin and the promoter segments were present in the construct. Exconjugants showing expected amplicons for both reactions were used in the next step and were considered to have all pieces of the terpentecin biosynthetic gene cluster construct, giving rise to *S. lividans* TK24 ΔactΔred::pTERP_1.

Assessing transcription of terpentecin genes in *S. lividans* TK24 ΔactΔred::pTERP_1

To confirm the heterologous genes were being expressed, reverse transcription PCR (RT-PCR) was performed using *S. lividans* TK24 ΔactΔred::pTERP_1 clone 1 and *S. lividans*TK24 ΔactΔred::pJC003 clone 1 as a negative control. 1.2 mL of culture was taken from a 50 mL culture in lean medium (in a 250 mL Erlenmeyer flask) at 21.4, 48 and 96 h of growth and preserved in RNA*later*® (Life Technologies). Ribonucleic acid (RNA) was extracted using the RiboPure™RNA Purification Kit (Life Technologies) and ProtoScript® M-MuLV *Taq* RT-PCR Kit (NEB) was used to synthesize complementary DNA (cDNA) according to manufacturers' instructions. A negative control with no reverse transcriptase was included. PCR was then performed on the cDNA using EconoTaq® and four sets of primers (Table 2.2): Terp_ORF9_F and 1d8R to amplify cDNA from ORF9 transcripts, 1d7F and 1d3R to target the terpentecin cyclase gene (ORF 11) cDNA, TERP_GGPPS_F and TERP_GGPPS_R to amplify GGPP synthase cDNA, TERP_HMG_coA_Red_F and TERP_HMG_coA_Red_R were used to amplify cDNA from the 3-hydroxy-3-methyl-glutaryl-CoA (HMG-CoA) reductase gene transcripts. The *hrdB* gene, which is constitutively transcribed and encodes a vegetative sigma factor, was used as a positive control using HrdB_F and HrdB_R primers [71] (Table 2.2). PCR cycling conditions are listed in Table 2.4.

Assessing terpentecin production in *S. lividans* TK24 ΔactΔred::pTERP_1

To confirm the production of terpentecin in *S. lividans* TK24 ΔactΔred::pTERP_1, seven recombinant colonies (1, 2, 9, 11, 40, 46 and 49) were cultured and the crude extracts were compared to that from three colonies (1, 2 and 4) of the control strain *S. lividans* TK24 ΔactΔred::pJC003 and the native terpentecin producer, *K. griseola* MF730-N6; media blanks were incubated simultaneously. Two sequential 10 mL seed cultures in seed medium were

grown at 30°C for 48 h in 25 X 150 mm glass culture tubes at 200 rpm. Using a 5% inoculum of seed #2, the *S. lividans* TK24 ΔactΔred::pTERP_1 recombinant strains and control strains were grown in 10 mL of both lean and ISP2 media for 7 d at 30°C and 200 rpm. The cultures were extracted twice with 10 mL and 7 mL ethyl acetate (EtOAc) and shaking for one hour for each extraction. The organic layers were combined and evaporated under air and the crude extracts were submitted to LC-HRMS at a concentration of 0.5 mg/mL.

Samples were initially analysed by visually comparing ELSD and MS chromatograms of the recombinant strain and the control strains to see if there were any obvious differences in peaks present. Next, mass chromatograms were monitored for ions with a mass to charge ratio (*m/z*) of 365.1959 (±5 ppm), consistent with the predicted $[M+H]^+$ pseudomolecular ion of terpentecin. Finally, a metabolomic analysis, as described in Forner et al.[24] and also in Chapter 6, was performed to highlight ions which differentiated the crude extracts of *S. lividans* TK24 ΔactΔred::pTERP_1 strains from that of the control strains. Intensity threshold was set to 1E3 and data were converted to a binary format.

2.2.4 Generation of two knock-out strains: *K. griseola* Δbaf and *K. griseola* Δterp

In order to better understand the metabolome of *K. griseola* MF730-N6 and to remove any confounding factors in the antibacterial assay against *B. subtilis* (described below), two strains were developed: one deficient in the production of the polyketide bafilomycin (*K. griseola* Δbaf) and the other deficient in the production of the diterpene terpentecin (*K. griseola* Δterp). The arrangement of the bafilomycin gene cluster (Figure 2.8) spans over 70 kb in both *S. lohii* and *K. setae*. To this end, most of the bafilomycin genes and a small but essential portion of the terpentecin biosynthetic gene cluster was removed and the apramycin resistance gene inserted via double crossover of homologous arms.

Construction of pBAFKO and pTERPKO

Two regions flanking the regions to be deleted were PCR amplified from gDNA using primers pairs bafKO_left_F/bafKO_left_R, bafKO_right_F/bafKO_right_R, terpKO_left_F/terpKO_left_R and finally terpKO_right_F/terpKO_right_R (Table 2.2)

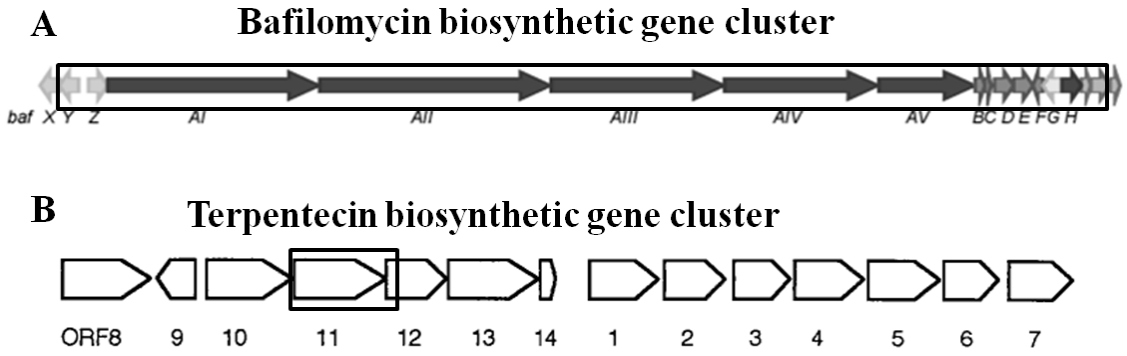


Figure 2.8. Arrangement of the bafilomycin **(A)** and terpentecin **(B)** biosynthetic gene clusters modified from Zhang et al.[77] and Dairi et al.[64]; the regions removed from the genome are shown in black boxes.

according to the conditions listed in Table 2.4 to give rise to amplicons bafKO_left (8,335,5343 bp-8,336,661 bp = 1118 bp), bafKO_right (8,407,725 bp-8,408,740 bp = 1015 bp), terpKO_left (12,000 bp-13,025 bp = 1025 bp) and terpKO_right (14,601 bp-15,584 bp = 983 bp), respectively. Following column or gel-purification of all arms, the left arms (bafKO_left and terpKO_left) were restricted with EcoRI-HF and SphI-HI, while right arms were restricted with HindIII-HF and SphI-HF to generate the proper sticky ends required for the assembly of the deletion plasmids. In doing so, the terpentecin left homologous arm was reduced to region 12,877-13,025 = 148 bp while the right arm covered region 15,397-15,584 = 187 bp.

The apramycin resistance gene (*aac(3)IV*) was used as a selective marker in the final knock-out plasmid constructs and was amplified from pJV177 (unpublished, L.C. Vining) using the primers pJV177_aprR_F/pJV177_aprR_R (Table 2.2) according to the conditions listed in Table 2.4. The 977 bp *aac(3)IV* amplicon was gel-purified and restricted with SphI-HF. Finally, the vector pSET151[69] was restricted with HindIII-HF and EcoRI-HF and the 6 kb band was gel-purified.

All components of the deletion plasmids (left arm, right arm, *aac(3)IV* and the plasmid pSET151) were ligated using T4 DNA ligase. Five microliters of each ligation reaction was used to transform chemically competent *E. coli* 10β and plated on LB containing 50 µg/mL apramycin. Colony PCR was performed on transformants harbouring pBAFKO using primers bafKO_left_F/bafKO_right_R (Table 2.2) using the conditions listed in Table 2.4 to generate a 3109 bp amplicon. Plasmid DNA was extracted from clones and restricted using SacI/PstI and AatII for the bafilomycin knock-out plasmid and HindIII and SacI /AatII for the terpentecin knock-out plasmid.

<u>*In vivo* deletion of the bafilomycin and terpentecin gene clusters</u>

The knock-out constructs, pBAFKO and pTERPKO, were used to transform chemically competent *E. coli* ET12567::pUZ8002 and plated on LB containing 50 µg/mL apramycin. The plasmids (pTERPKO and pBAFKO) were then introduced into *K. griseola* MF730-N6 via

conjugation with the aforementioned *E. coli* strains using the method of Choi et al.[78], which differs from the *Streptomyces* method of Keiser et al.[71] by omitting the heat-shock treatment for the spores and using ISP4 containing 30 mM $MgCl_2$ as the plating medium. After 16 h of growth at 30 °C, plates were overlaid with 10 µg/mL nalidixic acid and 25 µg/mL apramycin.

The apramycin resistant colonies were patched onto mannitol-soy medium containing 50 µg/mL thiostrepton and colonies that were resistant to apramycin but sensitive to thiostrepton, indicative of a double crossover and therefore gene deletion, were screened by colony PCR as described previously. PCR reactions to confirm the genotype of *K. griseola* Δbaf were performed as listed in Table 2.4 with primers bafKO_left_F, bafKO_right_R, and baf_Z_F and baf_Z_R (Table 2.2).

To confirm the genotype of *K. griseola* Δterp, PCR was performed on extracted gDNA using primers SEQTERP-1-d5-F and SEQTERP-2-d5-R (Table 2.2) and the cycling conditions listed in Table 2.4. Amplicons from terpentecin knock-out strains were also sequenced with primers SEQTERP-1-d5-F, SEQTERP-2-d5-R, TERP_602_F and TERP_1660_R to further confirm the genotype.

Assessing bafilomycin and terpentecin production in knock-out strains

Putative knock-out strains with confirmed genotypes were cultured in duplicate in 10 mL of lean medium along with the wild-type *K. griseola* MF730-N6 in order to characterize the phenotype of each strain. Cultures were grown for 48 h at 30°C and 200 rpm. The fermentations were extracted with EtOAc as described previously and analysed by LC-HRMS. When necessary, a metabolomic analysis was conducted in order to reveal differences between metabolomes by adapting the methods of Forner et al.[24] using a threshold value of 1E3. Once exported to Microsoft® Excel® (Redmond, USA), ions present in wild-type strains and absent from all replicates of the knock-out strains were selected. A principal component analysis (PCA), with no binary transformation of the data, was also used to determine which ions differed most between *K. griseola* MF730-N6 (WT) and *K. griseola* Δterp (KO) extracts.

Fermentation broths and crude extracts were also tested for inhibitory activity against *Bacillus subtilis* NRRL-B558 in an agar plate disk diffusion assay where *B. subtilis* from an overnight LB liquid culture was plated on LB agar using sterile cotton swabs. Disks (6 mm) containing the solution to be tested, either 20 µL fermentation broth or 20 µL of a 0.5 mg/mL crude extract in MeOH, was first dried and placed on the inoculated agar plate. 20 µL MeOH and 20 µL from media blank fermentations were used as negative controls. Plates were incubated overnight at 30 °C and inhibition of growth was evidenced by zones of inhibition.

2.2.5 Reassessing terpentecin production in *S. lividans* TK24 ΔactΔred::pTERP_1.

Three recombinant *S. lividans* TK24 ΔactΔred::pTERP_1 (colonies 1, 2 and 3) and three negative controls strains *S. lividans* TK24 ΔactΔred::pJC003 (colonies 1, 2, and 3) were cultured twice in 10 mL seed medium using 5% innocula in 150 x 25 mm culture tubes for 48 h and then in 10 mL medium at 30°C and 200 rpm for 3 d. The fermentation media used were MYM, tryptic soy broth (TSB), yeast extract-malt extract (YEME) +34% sucrose+0.5% glycine, lean medium, International *Streptomyces* Project medium 3 (ISP3) and mannitol-soy; their components are listed in Table 2.3. The fermentations were extracted, analysed by LC-HRMS and subjected to metabolomic and statistical analyses using binary transformed data as described previously for the knock-out experiments.

2.3 RESULTS AND DISCUSSION

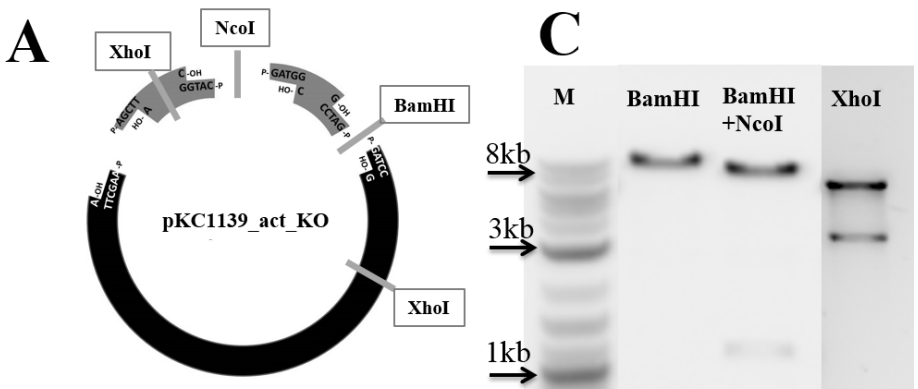
2.3.1 Construction of *S. lividans* TK24 ΔactΔred

To eliminate actinorhodin biosynthesis in *S. lividans* TK24 ΔactΔred, it was not necessary to delete the entire gene cluster as the compound is synthesized by a type II PKS which uses the core polyketide enzymes (acyltransferase, ketosynthase and acyl carrier protein) iteratively in order to assemble the polyketide backbone; the removal of any of the core enzymes terminates actinorhodin production. Five essential genes for the production of actinorhodin were thus removed from the genome of *S. lividans* TK24, *act I ORF1-3, actVII* and *actIV*; these genes encoded for the actinorhodin β-ketoacyl synthase subunits α and β, acyl carrier protein,

bifunctional cyclase/dehydratase and cyclase, respectively. By amplifying 1 kb flanking regions and cloning into the HindIII and BamHI site of pKC1139, pKC1139_act_KO was generated and its proper assembly was confirmed by restriction analysis. This plasmid contained the apramycin resistance gene for selection, an *E. coli* replication function derived from pUC, an origin of transfer that allowed transfer of the plasmid from *E. coli* to streptomycetes and finally a *Streptomyces ghanaensis* derived replication function which is functional in *Streptomyces* only at temperatures below 34°C[69].

Figure 2.9 illustrates the banding pattern of pDNA from four colonies that had shown an amplicon of the correct size during the colony PCR screen. Only clone 16 showed the expected DNA fragments when digested with BamHI (8500 bp), BamHI/NcoI (7500 bp and 1000 bp) and NcoI/XhoI (5500 bp and 2500 bp); pDNA from clone 16 was, thus, used to transform competent *E. coli* ET12567(pUZ8002). The *E. coli* strain was chosen due to its inability to methylate DNA and thus it permitted the transfer of the DNA into *Streptomyces* that have a strong methylation specific restriction system while the non-transmissible shuttle plasmid pUZ8002 provided the transfer function required for the plasmid to be donated during conjugation[70].

Similarly, the proper assembly of pKC1139_red_KO was confirmed by subjecting pDNA from four colonies showing the 2500 bp band from colony PCR to restriction digest analysis using EcoRV, XhoI, KpnI and PstI. All four colonies showed the same correct banding pattern (Figure 2.10) and pDNA from clone 12 was used to transform competent *E. coli* ET12567(pUZ8002). Since a silent gene deletion method was employed, where no antibiotic resistance gene remained in the genome of the knock-out strains, selecting for a single crossover event was made possible due to the inability of the plasmid to replicate at temperatures above 34°C. Colonies that were grown at 37°C and were resistant to apramycin must have had undergone a single crossover where the plasmid had integrated to the genome at the *attB* site. Selecting for double crossover colonies was possible by sub-culturing single crossover colonies in medium without any antibiotic.

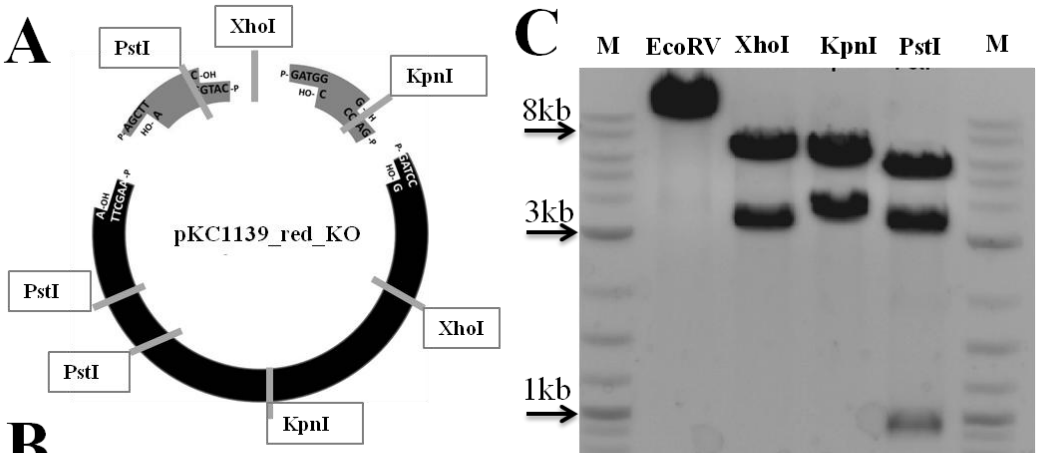


B

Expected band sizes

BamHI	BamHI +NcoI	XhoI
8500	7417	5811
	1083	2681

Figure 2.9. Restriction map of pKC1139_act_KO **(A),** the expected band sizes for each reaction **(B)** and composite agarose gel electrophoresis of restricted pDNA from pKC1139_act_KO clone sixteen showing the expected sized bands, M: 2-log ladder**(C)**.



B

Expected band sizes

EcoRV	XhoI	KpnI	PstI
9500	6280	6063	5435
	3197	3437	3195
			870

Figure 2.10. Restriction map of pKC1139_red_KO **(A),** the expected band sizes for each enzymatic reaction **(B)** agarose gel electrophoresis of restricted pDNA from pKC1139_red_KO clone 12 showing the expected sized bands, M: 2-log ladder **(C)**.

Genetic characterization

Exconjugants were subjected to genetic analysis to determine whether or not the desired *act*$^-$ genotype was achieved and to discriminate between double crossover and revertant colonies. Six of the investigated colonies had undergone gene deletion of the targeted actinorhodin genes as evidenced in Figure 2.11 by an amplicon of approximately 150 bp in colonies 18.40, 18.49, 18.74, 18.75, 18.76 and 6.1 when using primers annealing just outside the deleted region, confirming that the five genes were no longer present within the genome of these strains. In contrast, five colonies 18.14, 18.23, 18.52, 5.8 and 12.58 had an amplicon size around 5500 bp when using the same primer pair, consistent with a genotype where all five act genes were still present; these were colonies that had reverted back to the original genotype. The same 5.5 kb band was observed in the positive control using DNA of the wild-type organism *S. lividans* TK24.

Genetic analysis was also performed on exconjugants that arose from conjugation between *S. lividans* TK24 Δact clone 6.1 and *E. coli*::pKC1139_red_KO. From the 9 apramycin-sensitive red knock-out exconjugants tested, 5 colonies showed the expected 3042 bp band consistent with a *red*$^-$ genotype where 29 kb of the red biosynthetic gene cluster was removed (colonies14.31, 14.35, 4.9, 4.13 and 4.14). It was not possible to obtain amplicons from revertant colonies using the same primer pair since an amplicon length of 29 kb exceeded the length achievable by PCR. However, since a temperature sensitive origin of replication was used, only a double crossover event would have resulted in apramycin sensitive colonies with a 3 kb amplicon and thus no single crossover colonies remained.

Phenotypic characterization

The phenotype of various colonies was assessed based firstly on the presence of pigmented compounds in the fermentation medium (Figure 2.12). The culture of the wild-type strain TK24 was dark purple in colour, while the act knock-out strain 6.1 was red. The culture of strain 4.9, despite the *act*$^-$*red*$^-$ genotype, was light pink and finally the cultures of the remaining strains were beige, which was the colour of the MYM medium. Since only one PCR reaction

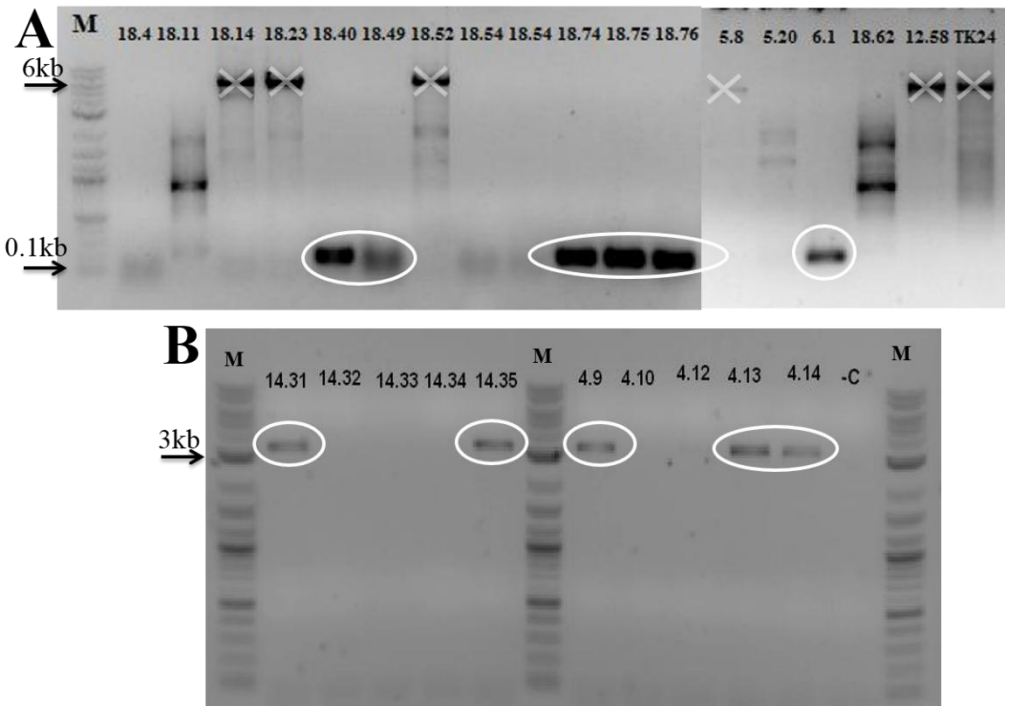


Figure 2.11. Gel image following colony PCR and electrophoresis to identify knock-out strains for actinorhodin **(A)** and undecylprodigiosin **(B)**; amplicons consistent with knock-out strains are circled in white and those consistent with wild-type stains are crossed-out, M: 2-log ladder.

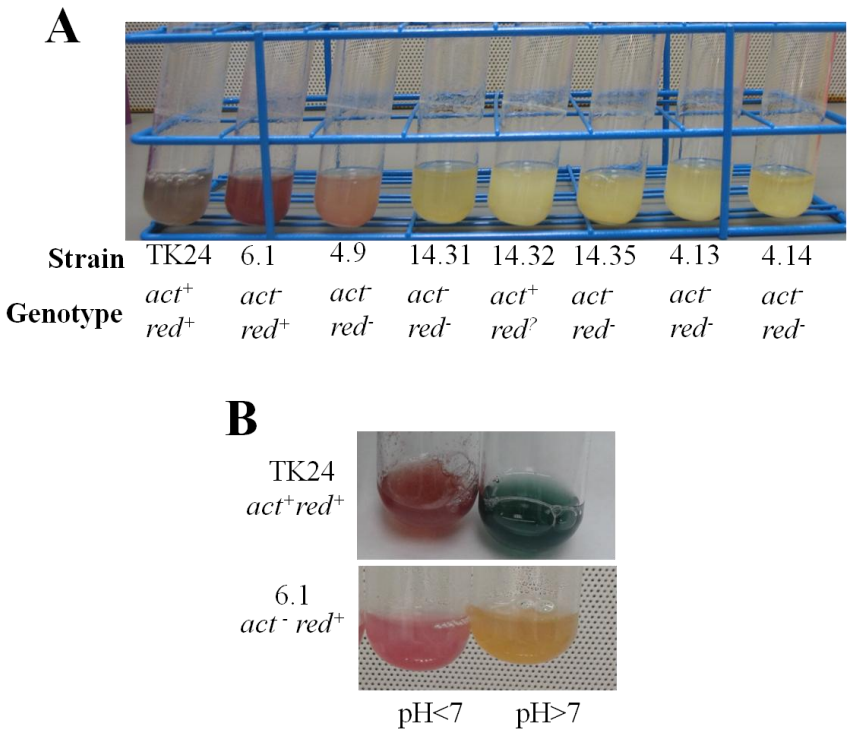


Figure 2.12. Assessment of pigmentation in various act and red knock-out strain of *S. lividans* TK24 according to the genotype assessed previously **(A)** and colour changes caused by the presence of actinorhodin and undecylprodigiosin in *S. lividans* TK24 and the *S. lividans* TK24 Δact, respectively, under acidic and basic conditions **(B)**.

was used to determine the genotype of the red knock-out strain in combination with apramycin sensitivity, it is possible that the strain 4.9 had not undergone double crossover. It is possible that the single crossover strain had lost resistance to the antibiotic via another recombination event or via a sudden deleterious mutation in the origin of replication or that the colony was contaminated with wild-type cells. Consequently, this strain was not used in future experiments. A more appropriate genotyping experiment would have involved another set of primers that would have amplified a region nested within the deleted genes. Phenotypic characterization of the strains was able to discriminate between the two undesirable compounds since actinorhodin was known to be red under acidic conditions and blue in basic conditions, while undecylprodigiosin was known to be red under acidic conditions and yellow under basic conditions[79, 80]. As illustrated in Figure 2.12 (B), the wild-type strain (*S. lividans* TK24) contained mostly actinorhodin, evidenced by the red and blue colouration of the fermentation when HCl and NaOH were added, respectively. However, when no actinorhodin was produced in *S. lividans* TK24 Δact (6.1), the presence of the compound undecylprodigiosin could be detected, evidenced by the appropriate colour change from red to yellow in response to NaOH addition. No colour change was observed in the double knock-out strain *S. lividans* TK24 ΔactΔred (clone 14.35).

The amount of act measured spectrophotometrically in the cultures is shown in Figure 2.13 (A). Act concentrations of strains 6.1, 18.4, 18.74, 18.75 and 18.78 varied between 0.4 and 5.5 µM and were significantly lower than the wild-type and the revertant strain 18.23 which showed average act concentrations of 37.7 and 32.1 µM, respectively; $F=1356.646$, $p<0.0001$. Previously reported act concentrations using this technique were approximately 20 µM in *Streptomyces coelicolor* A3(2)[74].

The amount of undecylprodigiosin measured in cultures of strains 4.13, 4.14, 14.31 and 14.35 (Figure 2.13 (B)) varied between 0.6 and 1.2 µM and were significantly lower than that from the wild-type and strains 6.1 and 4.9 which showed average red concentrations of 2.7, 6.4

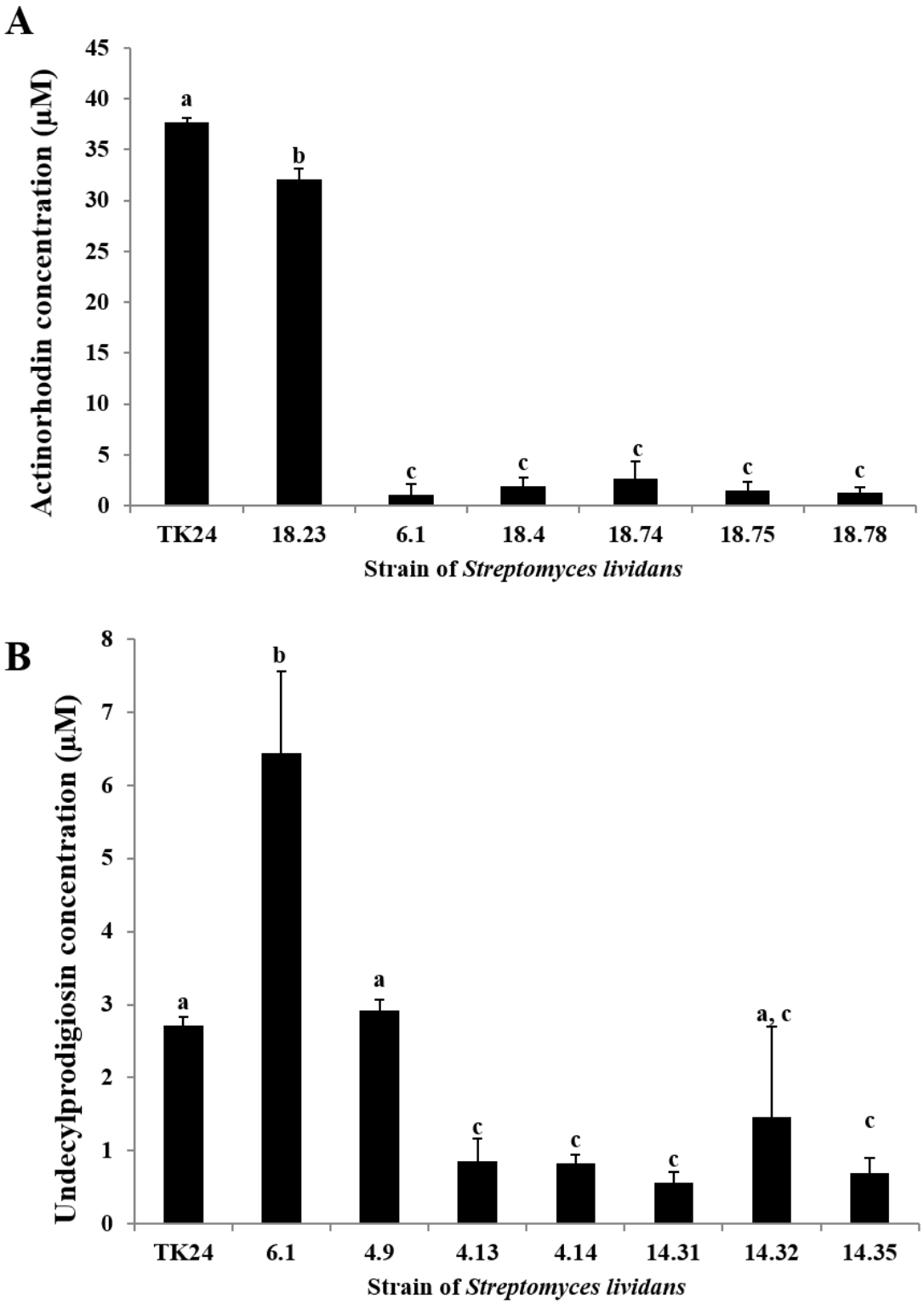


Figure 2.13. Amounts of compound quantified spectrophotometrically for various strains of the act knock-out **(A)** and red knock-out **(B)** procedures and shown as mean+1 standard deviation, letters represent groups that are significantly different from one another in the ANOVA using α=0.05, n=5 and n=3, respectively.

and 2.9 µM, respectively; F=31.8519 p<0.0001. The analysis also revealed that the act knock-out strain 6.1 contained significantly higher amounts of red than the wild-type strain, perhaps due to the lack of competition for acetate units necessary for the biosynthesis of both act and red[72, 73]. Previously reported red concentrations using this technique were of approximately 2 µM in *Streptomyces coelicolor* A3(2)[74].

Since this technique was based on spectrophotometric measurements and thus the transmission of light, it is likely that the readings observed in the knock-out strain were caused by background absorbance caused by other bacterial metabolites rather than by the presence of act and red. An alternative negative control could have been a closely related strain of *Streptomyces* incapable of producing act and red.

The residuals were tested for normality (Shapiro-Wilk W=0.927382; p=0.0234 and W=0.849277; p=0.0021 for act and red, respectively) and homoscedasticity (Bartlett's F=1.3204; p=0.2438 and F=3.1909; p=0.0022 for act and red, respectively). Although the assumption of normality and homoscedasticity were violated, the analyses were continued for two reasons: ANOVA analyses are robust[81] and slight changes in the p-value obtained due to a lack of normality would not have changed the conclusions drawn.

Due to consistency between the genotype and phenotype of strain 6.1, it was renamed *S. lividans* TK24 Δact and was used to generate the red knock-out strain. Similarly, strain 14.35 was renamed *S. lividans* TK24 ΔactΔred.

Such a host has been developed previously by Martinez et al. at Aventis Pharmaceuticals (now Sanofi) and its usefulness as a heterologous host was demonstrated by the production of granaticin, via the appearance of a purple compound and new peaks in the high performance liquid chromatography (HPLC) profile of the recombinant strain[82]. This was however not a commercially available strain and the development of such a strain would be beneficial for future uses in our laboratory. Although there are slight differences in the genotypes of the strain developed in this study and that which was previously published, where a smaller portion of the

act cluster was remove in this study, the phenotypes of both strains are the same: neither are capable of producing the endogenous natural products act and red. The strain *S. lividans* TK24 ΔactΔred developed in this study was thus used for the heterologous expression of the terpentecin biosynthetic gene cluster described below.

2.3.2 Construction of pTERP_1: The terpentecin biosynthetic gene cluster construct

Amplification of the terpentecin biosynthetic genes in three individual segments

To clone the terpentecin biosynthetic genes from *K. griseola* MF730-N6, PCR was used to amplify the gene cluster in three fragments: 6770, 2983 and 6558 bp. Each segment was successfully PCR amplified (Figure 2.14) and following gel-purification and restriction with the appropriate enzymes, each segment was successfully blunt-end cloned into pSET152.

Five transformants putatively containing cloned segment 1 (pSEGMENT1) were subjected to restriction analysis with the enzymes XbaI/NheI and NcoI (Figure 2.15). Clones number 1 and 5 contained plasmids consistent with a forward insertion of segment 1 at the EcoRV site of pSET152, while clones 2 and 4 showed a banding pattern consistent with the reverse incorporation of the insert. Clone 5 had extra bands when restricted with NcoI, this most likely due to partial digestion of the pDNA and clone 3 had a banding pattern inconsistent with the two acceptable options and therefore was not used.

When cloning blunt-ended DNA into a plasmid from which the genes will be expressed (expression plasmids), the direction of insertion is an important factor to consider ensuring correct transcription and translation of the inserted DNA. However, for the purposed of this research, the constructed plasmids' role were simply to provide sufficient quantities of each fragment for sequencing and to construct pTERP_1 and therefore the direction of insertion of the clone chosen was unimportant.

Similarly, Figure 2.16 illustrates that the pDNA from clones 1 and 2 of pSEGMENT2 showed restriction patterns consistent with inserts in the reverse direction while clone 5 was consistent with an insert in the forward direction in relation to the ΦC31 integrase gene of

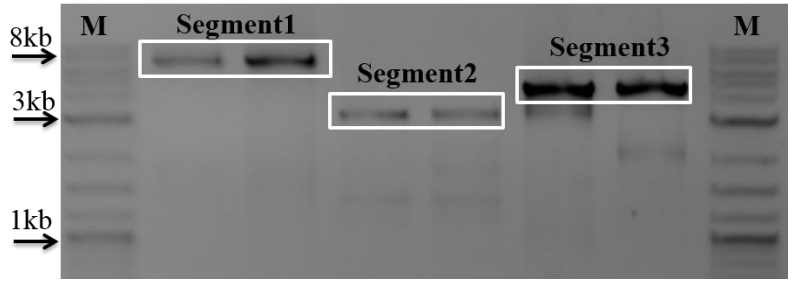


Figure 2.14. PCR amplicons of each terpentecin segment following gel electrophoresis; DNA that was gel-purified and cloned into pSET152 is shown in boxes, M: 2-log ladder.

A

Expected fragment sizes (forward)			
XbaI +NheI	NcoI	XhoI	BamHI
9091	8373	6114	6502
3402	2453	3176	5661
	1767	2572	301
		631	29

Expected fragment sizes (reverse)			
XbaI +NheI	NcoI	XhoI	BamHI
9091	5674	5311	5708
3402	5052	3979	5661
	1767	2572	823
		631	301

B

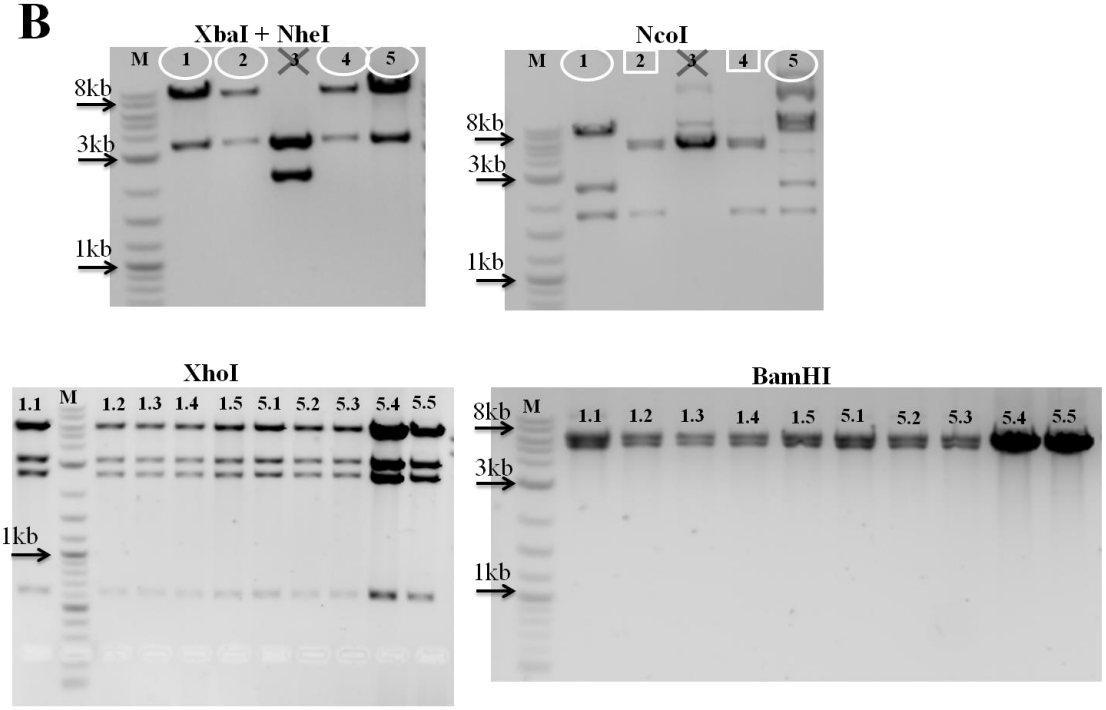


Figure 2.15. Predicted fragment sizes of pSEGMENT1 plasmid DNA digests according to the direction of insertion in relation to the plasmid backbone **(A)** and picture of the gel following electrophoresis for each digest for various clones. Five clones were initially screened (1-5) with XbaI/NheI and NcoI and clones showing the expected banding pattern were replated and five colonies were subjected to restriction by XhoI and BamHI to assess population homogeneity (X.1-X.5) (**B)**; samples with a banding pattern consistent with a forward and reverse insert are circled and boxed, respectively, and those inconsistent are crossed-out; when insertional direction cannot be determined, a correct banding pattern is circled by default, M: 2-log ladder.

A

Expected fragment sizes (forward)				Expected fragment sizes (reverse)			
XbaI +NheI	SacI	XhoI	BamHI	XbaI +NheI	SacI	XhoI	BamHI
5316	5527	4997	6846	5316	4874	5856	7503
3402	2440	3721	1824	3402	3093	2862	1167
	751		48		751		48

B

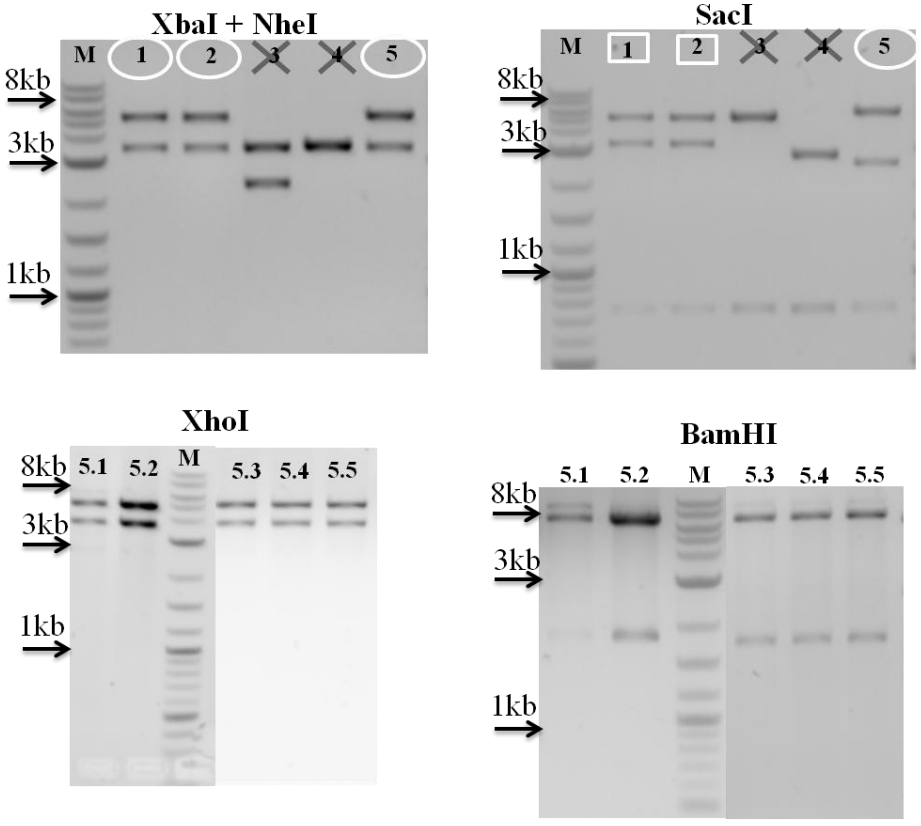


Figure 2.16. Predicted fragment sizes of pSEGMENT2 plasmid DNA digests according to the direction of insertion in relation to the plasmid backbone **(A)** and picture of the gel following electrophoresis for each digest. Five clones were initially screened (1-5) with XbaI/NheI and NcoI and clones showing the expected banding pattern were replated and five colonies were subjected to restriction by XhoI and BamHI to assess population homogeneity (X.1-X.5) **(B)**; samples with a banding pattern consistent with a forward and reverse insert are circled and boxed, respectively and those inconsistent are crossed-out; when insert direction cannot be determined, a correct banding pattern is circled by default, M: 2-log ladder.

pSET152. Clones 3 and 4 were discarded as their restriction patterns were not consistent with either direction and the sum of the fragment sizes were much smaller than expected (8718 bp).

Finally, restriction analysis of pSEGMENT3 (Figure 2.17) revealed that clones 1, 2, 4 and 5 had plasmids with inserts in the forward direction, while clone 3 had incorporated the insert in the opposite direction in pSET152. Although restriction analysis with XhoI and BamHI was performed on clones, 1, 2, 4 and 5, only one example is shown as they all showed the same pattern. Plasmid DNA from clone 2 was used for sequencing.

Plasmid DNA from pSEGMENT1-#5, pSEGMENT2-#5 and pSEGMENT3-#2 were sequenced and the assembly of the sequences with the reference sequences (AB048795.1 and AB037907.1) (Figure 2.18) revealed a total of 13 regions with inconsistencies between the segments and reference sequences, with segment 3 having the majority of inconsistencies. Fresh pDNA was resequenced and showed the same inconsistencies and thus ruled out sequencing errors. Ten PCR amplicons from *K. griseola* gDNA (Table 2.4) were sequenced at both strands for all regions with errors and the same discrepancies with the published terpentecin biosynthetic gene cluster were present in the genome, suggesting that these were not due to PCR generated errors in the cloned segments, but rather were probably sequencing errors present in the published sequence of the terpentecin biosynthetic gene clusters. With DNA sequencing technologies becoming faster, less expensive and more accurate through the discovery of better DNA polymerases, it is not surprising that this many errors were found in the published sequence. Furthermore, 10 of the 13 errors were present in the late part of the cluster, the mevalonate pathway genes, which were submitted to GenBank in 1998 (accession AB015626.1) while the terpentecin specific genes were submitted to GenBank in 2001 (accession AB048795.1). The terpentecin segments were thus proven to be error free and were ready to be assembled.

Construction of the constitutive promoter *ErmE*_actII-ORF4*_JC.

To remove the reliance on native promoters to drive the transcription of the terpentecin gene

A

Expected fragment sizes (forward)				Expected fragment sizes (reverse)			
XbaI +NheI	SacI	XhoI	BamHI	XbaI +NheI	SacI	XhoI	BamHI
5409	6613	6725	6362	7674	7364	7704	8326
3483	4930	5569	2647	3402	4179	4590	1595
3402	751		1595	1218	751		1383
			1386				683
			307				307

B

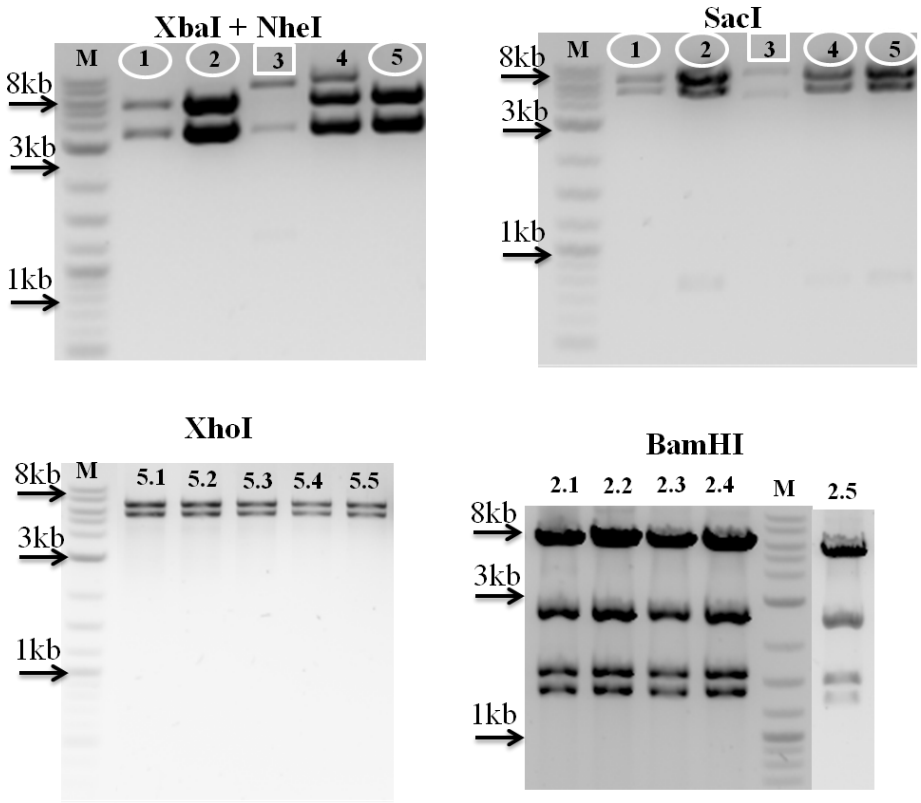


Figure 2.17. Predicted fragment sizes of pSEGMENT3 plasmid DNA digests according to the direction of insertion in relation to the plasmid backbone **(A)** and picture of the gel following electrophoresis for each digest. Five clones were initially screened (1-5) with XbaI/NheI and NcoI and clones showing the expected banding pattern were replated and five colonies were subjected to restriction by XhoI and BamHI to assess population homogeneity (X.1-X.5) **(B)**; samples with a banding pattern consistent with a forward and reverse insert are circled and boxed, respectively, M: 2-log ladder.

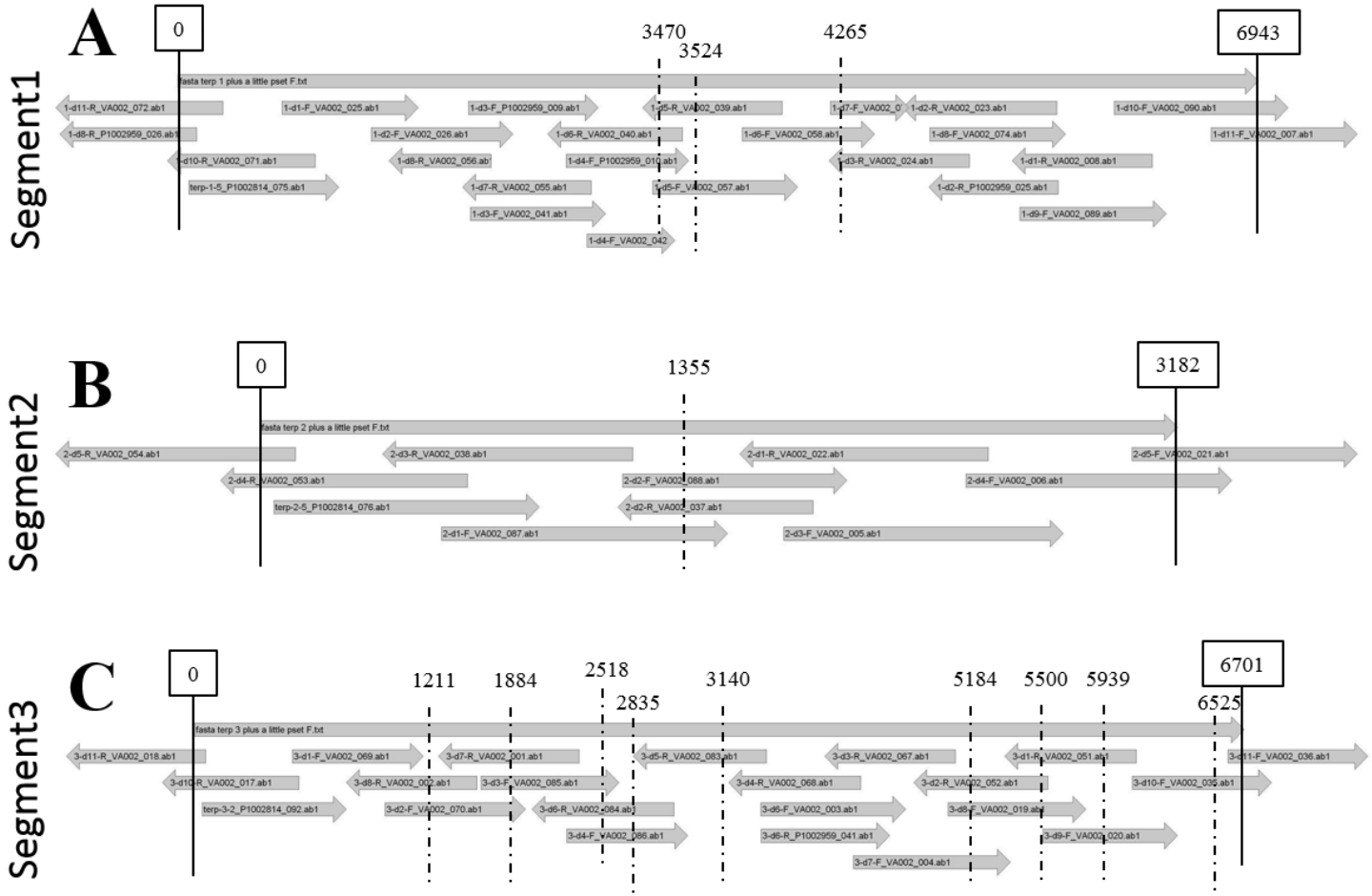


Figure 2.18. Assembly of reference sequence (top) with sequencing results of pSEGMENT1-#5 **(A)**, pSEGMENT2-#5 **(B)** and pSEGMENT3-#2 **(C)** illustrating the location of all 13 inconsistencies between the cloned and reference gene cluster sequences; boxed numbers represent start and end of each segment, while dashed lines represent the locations of inconsistencies.

cluster, the actinorhodin activator-promoter system described in Wilkinson et al.[76] was used in this construct. The constitutive *ErmE** promoter drives the transcription of *actII-ORF4*, whose protein then activates the *actI* promoter, which in turns ensures the transcription of the biosynthetic genes downstream of p*actI*. To this end, two fragments needed to be assembled.

The promoter fragment A was provided by Bio Basic Inc. within the SmaI site of pUC57 and showed the expected banding pattern (Figure 2.19 (A)). The analysis confirmed the presence of unique SalI and AgeI restriction sites and thus allowed the unidirectional insertion of fragment B at those sites. Fragment B was successfully amplified from *S. lividans* TK24 gDNA (Figure 2.19 (B)) and cloned into the aforementioned sites. Following transformation of competent *E. coli* with the newly ligated reaction, the resulting fragments from digested pDNA of clone 4 (Figure 2.19 (D)) were consistent with the predicted fragment sizes. The promoter was sequenced using multiple primers and there were no discrepancies between the sequenced plasmid DNA and the promoter reference sequence; pUC57::P*ErmE*_actII-ORF4*_JC pDNA was isolated in high yield and prepared accordingly for pTERP_1 construction.

Construction of a new *Streptomyces* integrating cosmid vector: pJC003

Since *Streptomyces* integrating plasmid vectors such as pSET152 would not remain stable with a 14 kb insert, it was necessary to construct a cosmid vector that would allow integration of the DNA at the *attB* site of the *Streptomyces* genome. To this end a 4 kb fragment from pSET152 containing the apramycin resistance gene, the origin of transfer and the ΦC31 integrase and attachment site (*attP*)[69] was PCR amplified (Figure 2.20) and ligated to a 4 kb HpaI fragment from pWEB containing the ampicillin resistance gene, the *ColE1 E. coli* origin of replication and a bacteriophage *cos* site for λ packaging. Plasmid DNA from 5 transformants was digested with restriction enzymes (Figure 2.20); clones 1 and 4 showed a restriction pattern consistent with a forward insertion of the pWEB fragment, where the translational direction of all genes in both fragments was in the same direction. In contrast,

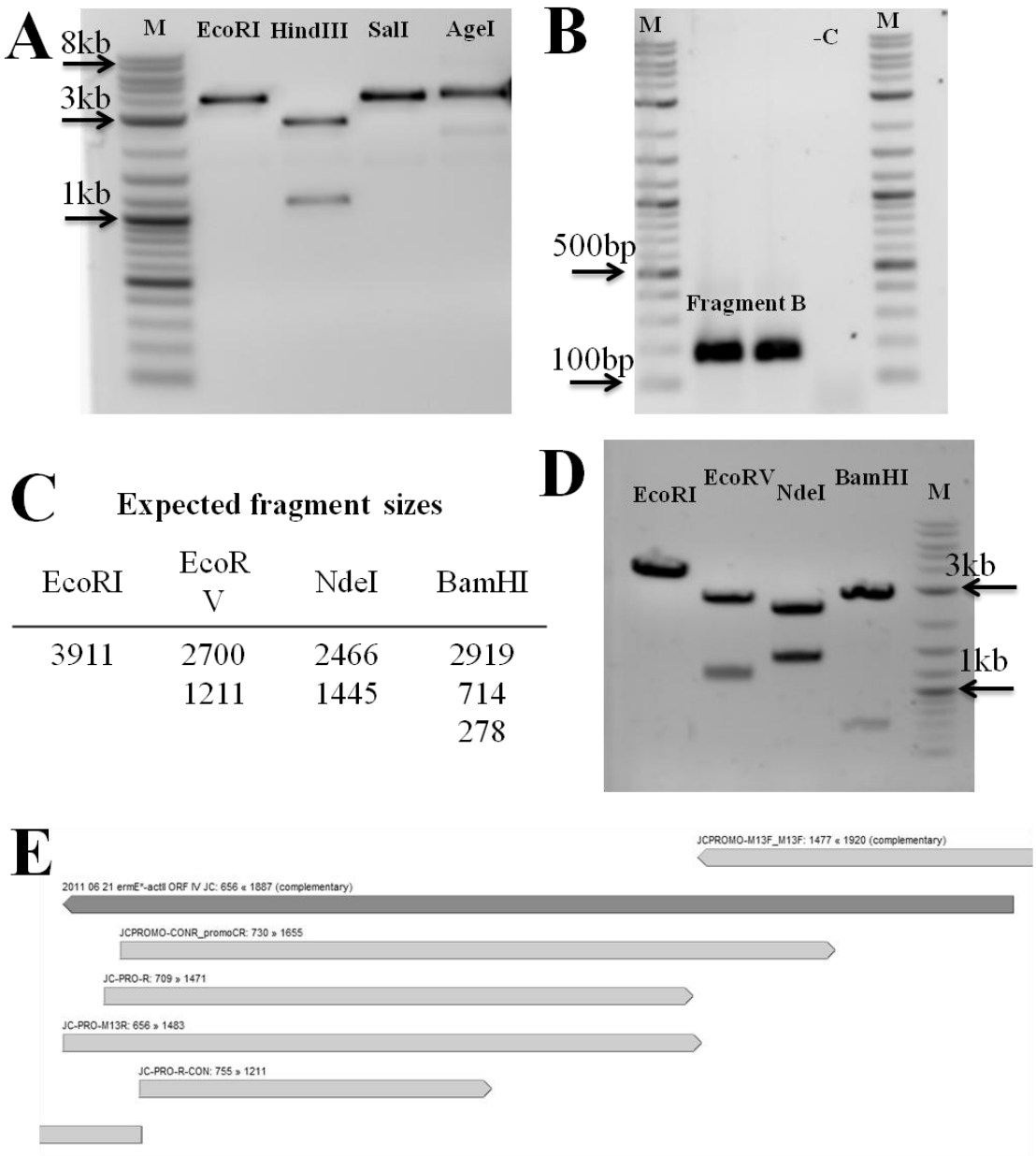


Figure 2.19. Restriction analysis of pUC57::fragment_A **(A)**, PCR amplification of fragment B from *S. lividans* TK24 gDNA **(B)**, banding pattern of plasmid DNA from pUC57::P*ErmE*_actII-ORF4*_JC expected **(C)** and observed from clone 4 **(D)** and overview of the assembly of sequencing results (light grey) with the promoter reference sequence (dark grey) confirming the proper assembly and accuracy of the constructed promoter sequence **(E)**, M: 2-log ladder.

Expected fragment sizes (forward)				Expected fragment sizes (reverse)			
EcoRV	NheI	HindIII	AclI	EcoRV	NheI	HindIII	AclI
8126	5602	4413	6257	8126	4334	5281	6257
	2524	2744	1496		3792	1876	1496
		969	373			969	373

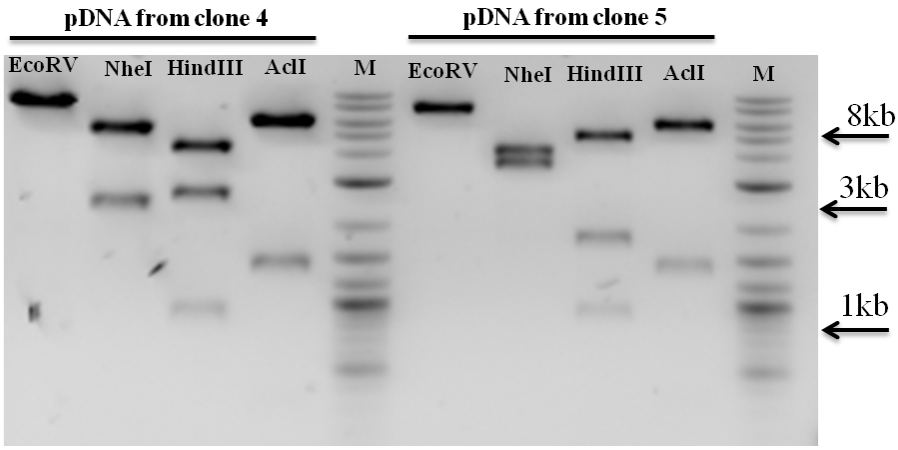


Figure 2.20. Predicted and observed DNA fragments following digestion and gel electrophoresis of pJC003 clones 4 and 5 illustrating a "forward" and "reverse insertion of the 4 kb pSET152 fragments, respectively, when comparing the translational direction of the genes from both fragments, M: 2-log ladder.

clones 2, 3 and 5 showed restriction patterns consistent with opposite translational directions. Conjugation efficiencies of the resulting shuttle vector constructs were tested to ensure their suitability for heterologous expression studies. Conjugation efficiencies observed for colonies 4 and 5 in *S. lividans* TK24 ΔactΔred were similar, where each construct resulted in a lawn of exconjugants, thus clone 4 was used as the vector in the construction of pTERP_1.

Assembly of the entire terpentecin biosynthetic gene cluster in pJC003 -pTERP_1

To construct pTERP_1, segments 1-3 of the terpentecin gene cluster were combined with the *ErmE*_actII-ORF4*_JC promoter and pJC003 in a 5-way ligation which was used to transform competent *E. coli.* Transformants (171) were screened by colony PCR for the presence of segments 2 and 3 of the terpentecin gene cluster. A single clone (#7) supported the amplification of the expected 3.9 kb fragment (Figure 2.21 (A)). Clone 7 was further screened for the presence of segment 1 and 2 and the expected 856 bp amplicon was obtained (Figure 2.21 (B)), giving rise to pTERP_1 (Figure 2.22).

The correct assembly of pTERP was further confirmed by sequencing the cloning junctions using a variety of primers and is summarized in Figure 2.23; the sequences matched the expected sequences exactly and indicated a reverse insertion into the vector when comparing the translational directions. To confirm the overall length of the construct, nine restriction digests of pTERP_1 were performed (Figure 2.24); all were consistent and confirmed the reverse insertion of the genes within the vector backbone. Due to the low transformation efficiency of the constructed plasmid, it was not possible to obtain the other conformation of the plasmid. The genetic evidence presented here supported the proper construction of pTERP_1 containing the terpentecin biosynthetic gene cluster under the control of a constitutive promoter.

2.3.3 Heterologous expression of terpentecin in *S. lividans* TK24 ΔactΔred

To heterologously express the terpentecin biosynthetic genes in *S. lividans* TK24 ΔactΔred, it was first necessary to transform *E. coli* ET12567::pUZ8002 with pTERP_1 and the

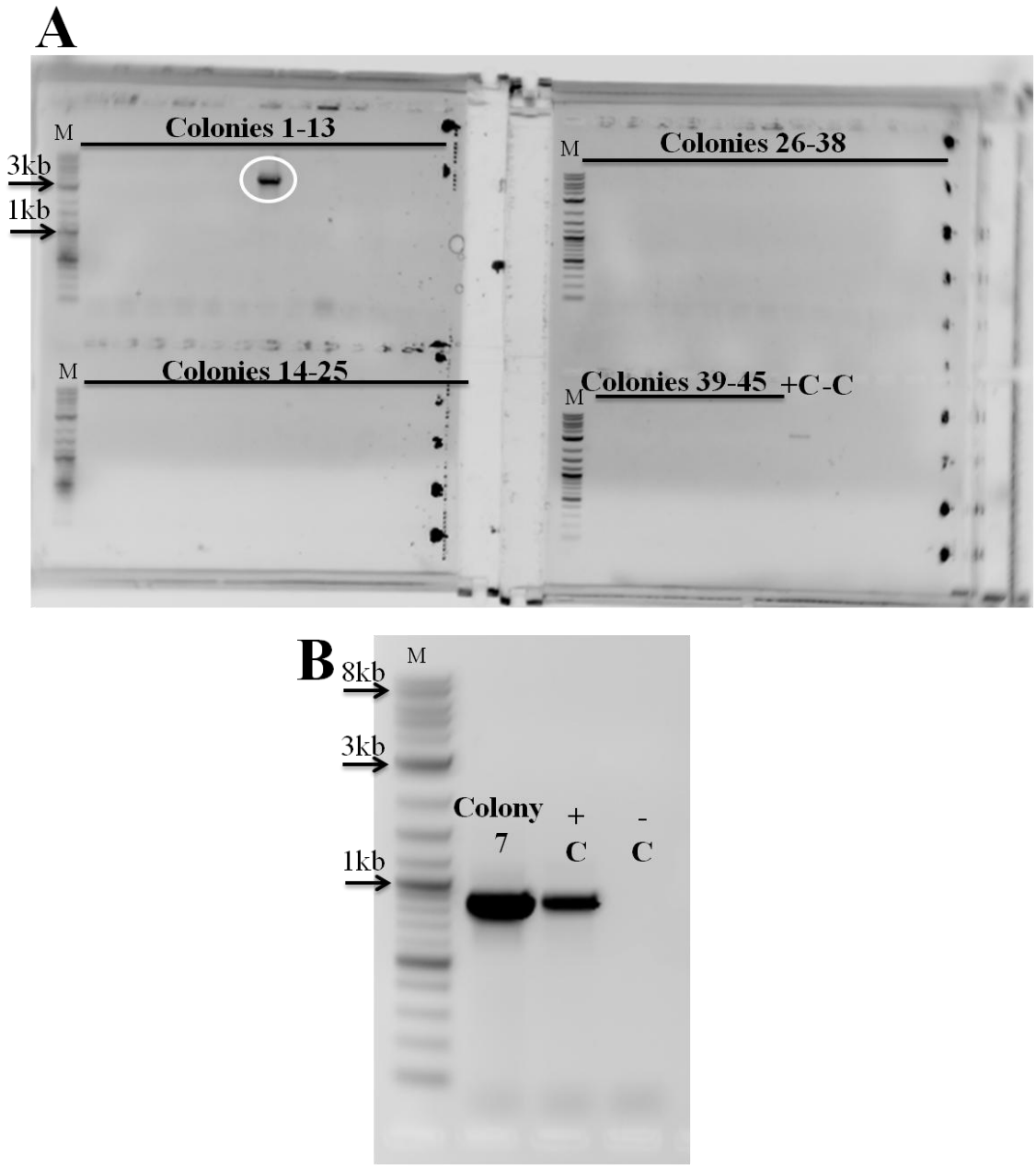


Figure 2.21. Colony PCR to screen transformants that contained terpentecin segments 2 and 3 **(A)** and colony PCR on clone 7 for DNA of terpentecin segments 1 and 2 **(B)**; *K. griseola* MF730-N6 was used as positive control (+C) and no DNA was added in PCR reaction for negative control (-C); clone #7 is circled, M: 2-log ladder.

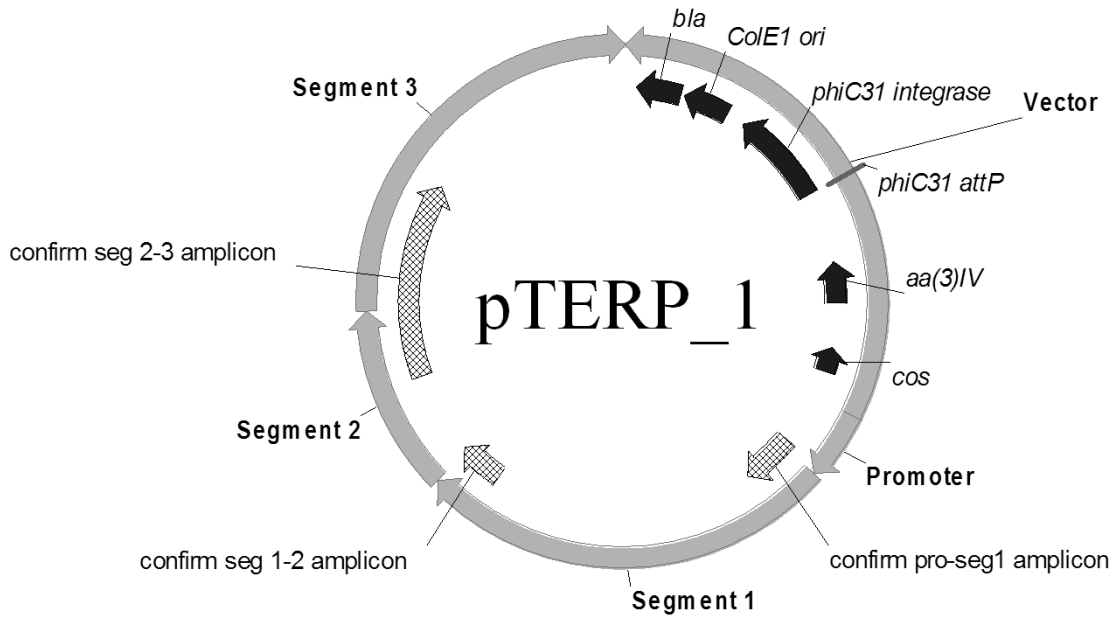


Figure 2.22. Map of pTERP_1 containing the terpentecin biosynthetic gene cluster under the control of the constitutive promoter *ErmE*_actII-ORF4*_JC. DNA fragments are shown in grey, amplicons used to screen transformants and exconjugants via colony PCR are shown in crosshatch and vector features are shown in black.

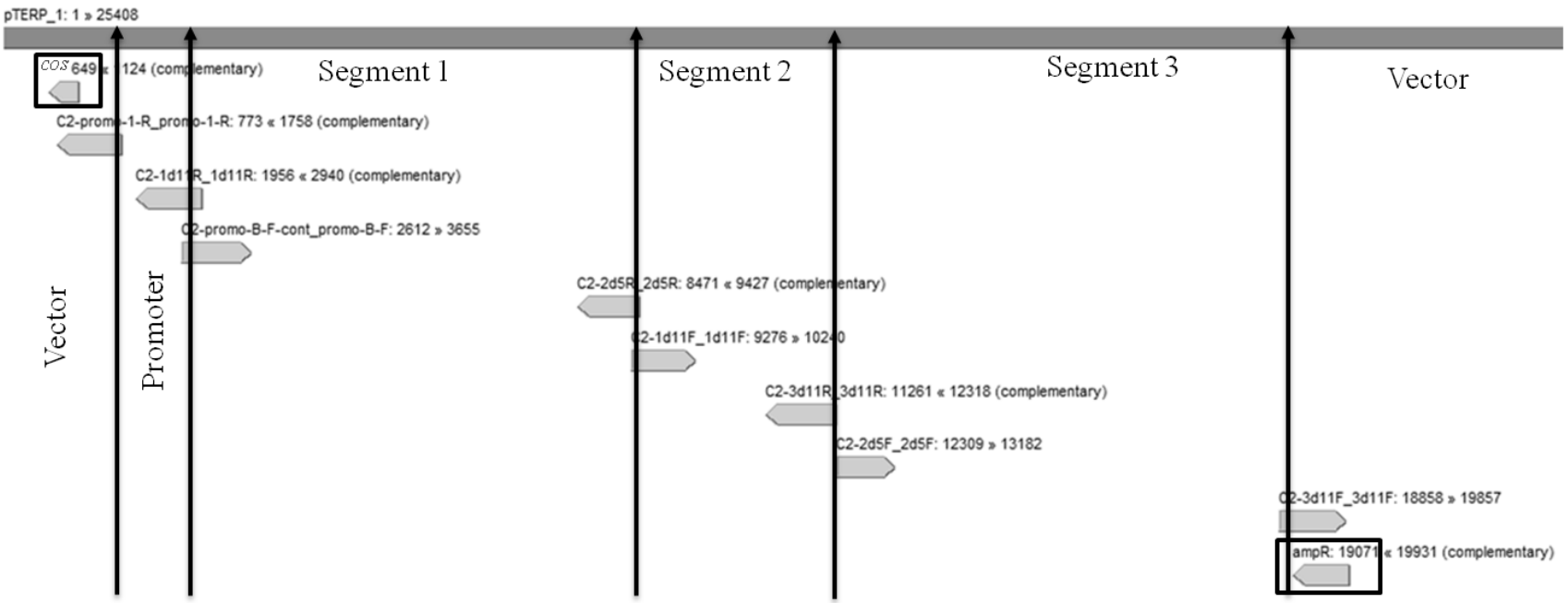


Figure 2.23. Overview of the assembly between the reference sequence for pTERP_1 (top, dark grey) and the sequencing results of the junctions from the resulting pDNA (light grey); the black arrows indicate the approximate location of the junctions and the *cos* and ampicillin resistance vector genes used to determine the direction of insertion are boxed.

A

Expected fragment sizes (forward)

1 BsrGI	2 NdeI +NheI	3 NsiI	4 NotI +NdeI	5 NcoI	6 AclI	7 SacI	8 HindIII	9 NsiI +NdeI
17637	16380	25436	11998	18116	11248	5905	17370	17550
4425	6532		7611	5553	7558	5569	4413	7886
3377	2524		2987	1767	6257	5394	2684	
			2840		373	4026	969	
						3791		
						751		

Expected fragment sizes (reverse)

1 BsrGI	2 NdeI +NheI	3 NsiI	4 NotI +NdeI	5 NcoI	6 AclI	7 SacI	8 HindIII	9 NsiI +NdeI
17637	21439	25436	11998	16293	11873	**9226**	19971	17550
4425	2524		7611	**7376**	6933	5905	4413	7886
3377	**1473**		2987	1767	6257	5763	969	
			2840		373	3791	83	
						751		

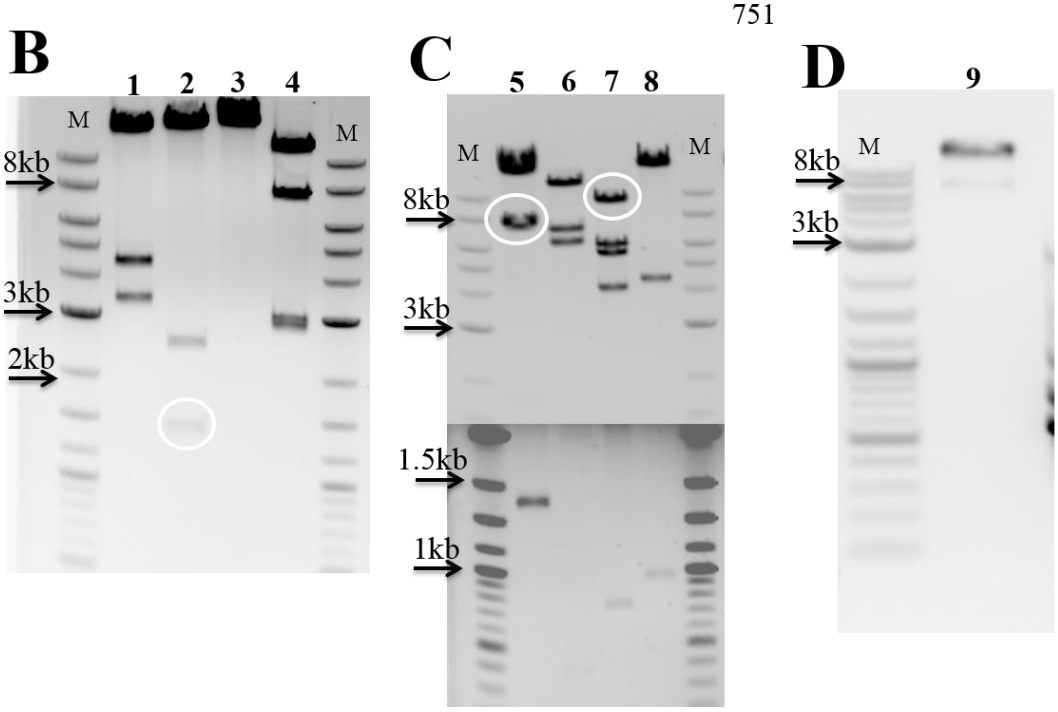


Figure 2.24. Predicted **(A)** and observed DNA fragments **(B-D)** from nine restriction digests of pTERP_1 clone 7; diagnostic bands are circled in white along with the expected sized fragment in bold, lanes are numbered according to the digest in A; gel C was taken in two separate pictures using different exposure times to allow the detection of smaller bands without oversaturation of the larger ones, M: 2-log ladder.

empty vector as a negative control strain. This ensured the foreign DNA would evade the native restriction system of *S. lividans* due to the mutations in dam, dcm and hsdM methyltransferases[83] of the *E. coli* strain, resulting in non-methylated DNA. The non-transmissible plasmid pUZ8002 also provided the transfer functions required during conjugation[70]. The non-methylating *E. coli* was then used as the donor strain during conjugation with *S. lividans* TK24 ΔactΔred.

Five pJC003 and 49 pTERP_1 apramycin-resistant exconjugants were preserved and chromosomal integration was confirmed by colony PCR of two regions within the terpentecin biosynthetic gene cluster (Figure 2.25): one amplicon covered the promoter to segment1 junction region (confirm_pro-seg1), while the other amplified the junction between segments 2 and 3 (confirm_seg2-seg3); both amplicons were present in recombinant strains whereas they were absent in control strains containing pJC003. All 49 clones showed the correct 1 kb amplicon spanning the promoter and the terpentecin segment 1, though some were faint. Colony PCR also determined that all clones other than 6, 7, 26 and 27 contained the 3.7 kb amplicon spanning terpentecin segments 2 and 3. PCR experiments could have been repeated using genomic DNA if necessary; numerous colonies were however shown to contain all fragments of the promoter and gene cluster.

Since the presence of the terpentecin biosynthetic genes was confirmed in 43 exconjugant colonies, reverse transcription RT-PCR was used to assess expression of key terpentecin biosynthetic genes. To this end, RNA was extracted from the recombinant strain, *S. lividans* TK24 ΔactΔred::pTERP_1 clone 1, and control strain, *S. lividans* TK24 ΔactΔred::pJC003 clone 1, following 21.5, 48 and 96 h of incubation in a 250 mL flask containing 50 mL of lean medium. The RNA was then converted to cDNA and used as a template to amplify various regions within the terpentecin biosynthetic gene cluster.

The results confirmed that various terpentecin biosynthetic genes throughout the cluster were successfully being expressed (Figure 2.26) as amplicons were obtained at all three time points for the recombinant strain, while being absent in the control host containing

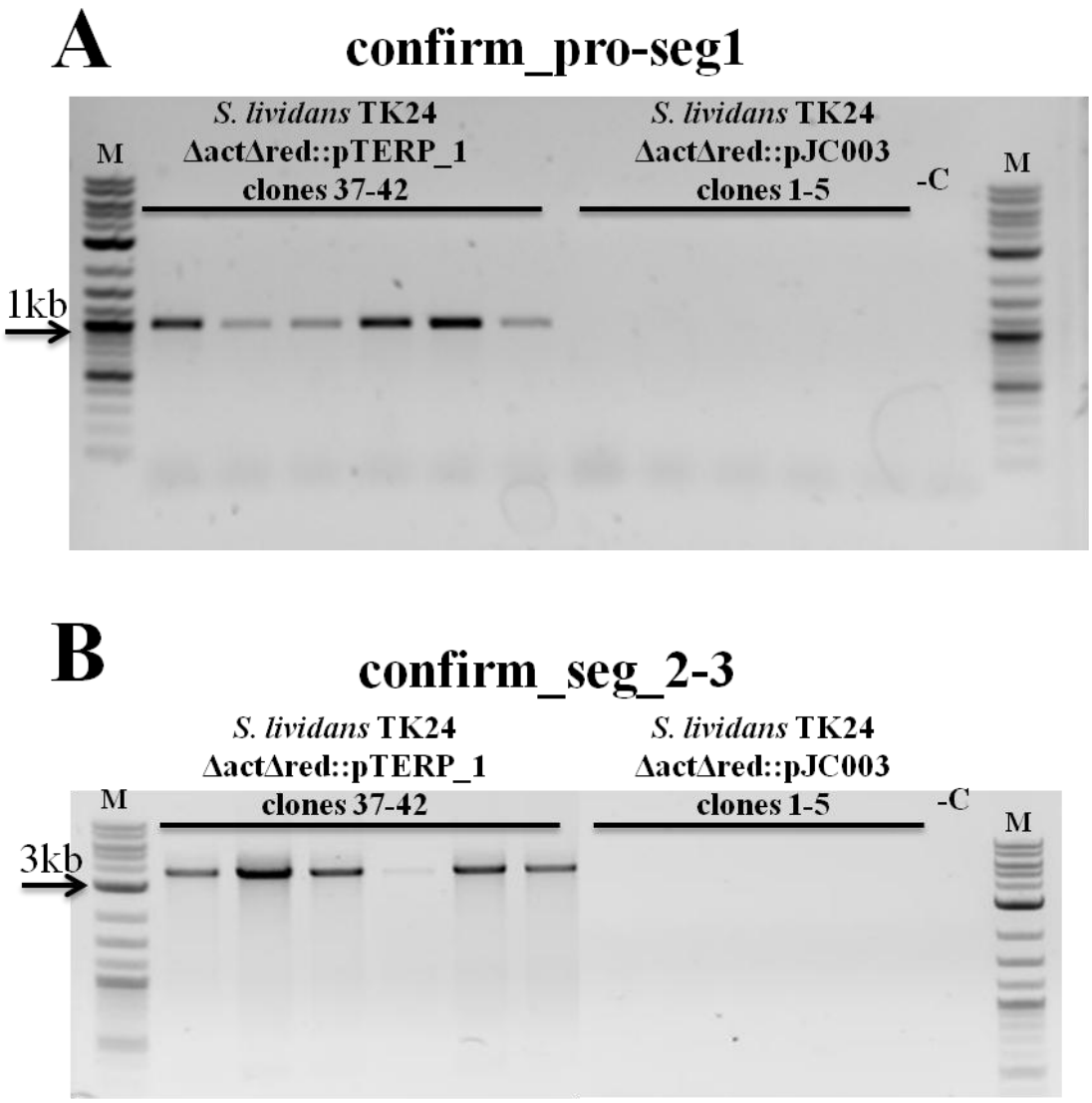


Figure 2.25. PCR screen for the presence of terpentecin genes by amplification of the region spanning the promoter to segment 1 **(A)** and segment 2 to segment 3 **(B)** regions in 49 clones of *S. lividans* TK24 ΔactΔred::pTERP_1; for simplicity, only the results from clones 37 to 42 as well as from the 5 control clones are shown along with a negative control (-C) where no DNA was added to the PCR reaction; M: 2-log ladder.

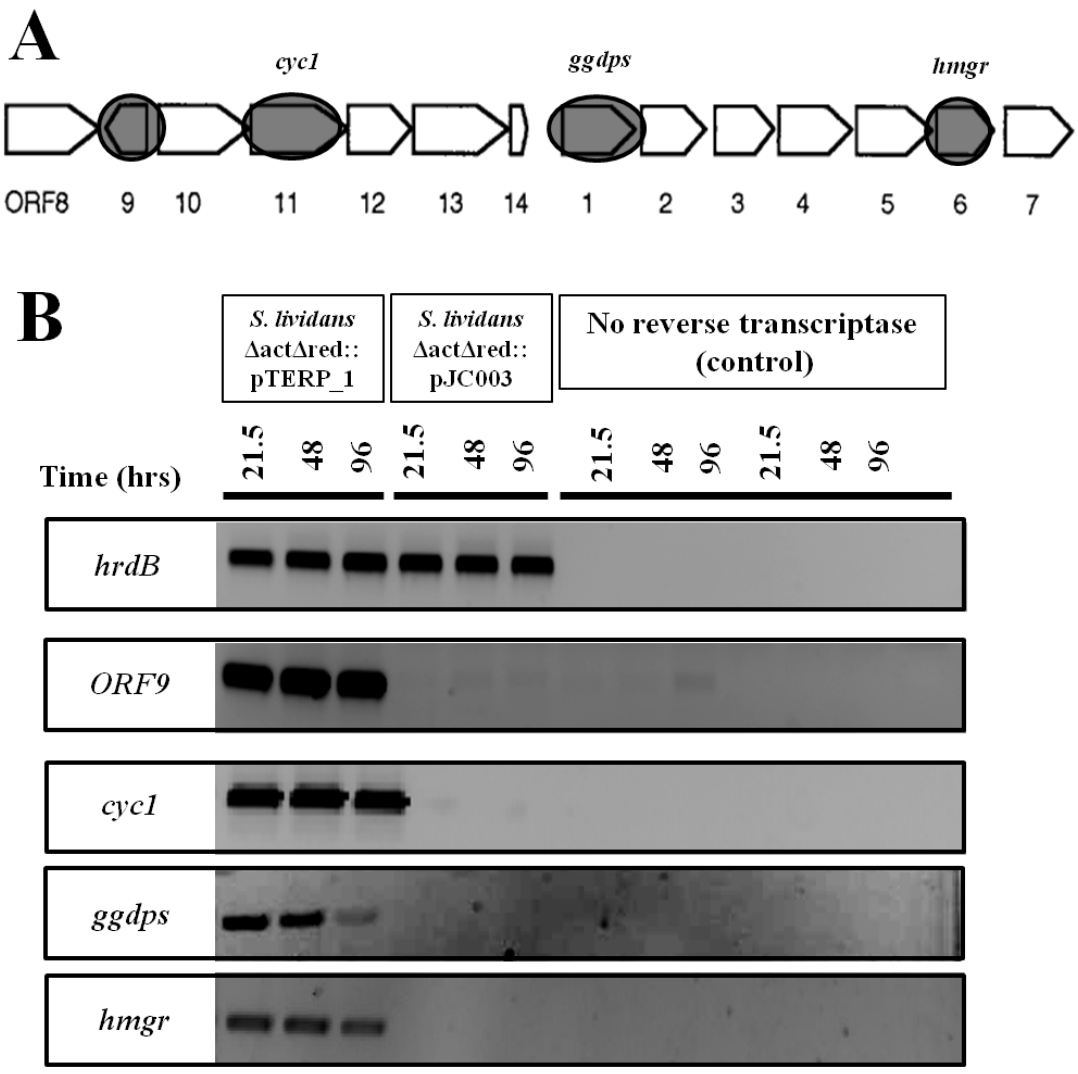


Figure 2.26. Monitoring the expression of terpentecin biosynthetic genes throughout the cluster; genes targeted in the RT-PCR experiment are circled **(A)**. Reverse transcription PCR experiment illustrating that heterologous terpentecin genes were being expressed in *S. lividans* TK24 ΔactΔred::pTERP_1 clone 1, while no expression was detected in *S. lividans* TK24 ΔactΔred::pJC00; agarose gel showing amplicons from cDNA of four terpentecin and one reference (*hrdB*) gene transcripts **(B)**; negative controls for each time point and strain were included to ensure amplicons were not a result of contaminating DNA (no reverse transcriptase enzyme was used during the protocol).

only the vector pJC003. It was also clear that amplicons originated from mRNA transcripts and not from residual genomic DNA as per the absence of amplicons in negative control reactions where no reverse transcriptase enzyme was used in the process. The positive control used here were mRNA transcripts from the *hrdB* gene, the major and essential sigma factor gene in *Streptomyces*, responsible for the transcription of housekeeping genes[84, 85]; this transcript was found in both recombinant and control strains at all sampling times (Figure 2.26).

The transcripts were detected early during the fermentation, as was expected due to the presence of the constitutive promoter *ErmE*_actII-ORF4*_JC directly upstream of the cluster. The terpentecin and mevalonate genes had previously been shown to be polycistronically transcribed, where *ORF 10 to 14* provided a 6.5 kb transcript and *ORF 1-7* provided a 8.5 kb transcript[64, 75]. PCR amplicons for *ORF 11* transcripts were thus necessary to prove that the first transcript was present, while monitoring of *ORF 1 and 6* proved that the latter transcript was present in the recombinant strain. Since *ORF 9* is transcribed from the complementary DNA strand, presence of the transcript was also monitored.

The evidence of transcription from *ORFs 9, 11, 1* and *6* shown here was thus sufficient to assess that the terpentecin biosynthetic genes were expressed in the recombinant strain. Due to the saturated bands resulting from the amplified cDNA following electrophoresis (Figure 2.26), it was not possible to quantitate the relative amounts of transcripts at different sampling times; earlier sampling times, reducing the amount of cDNA template or reducing the number of cycles during PCR may provide such information. However, the objective of the experiment was to assess whether or not the heterologous genes were expressed; the presence of such transcripts was sufficient to conclude that the terpentecin biosynthetic genes were being transcribed in the recombinant strain *S. lividans* TK24 ΔactΔred::pTERP_1 clone 1.

Assessing terpentecin production in *S. lividans* TK24 ΔactΔred::pTERP_1

Since the terpentecin biosynthetic genes were proven to be not only present, but to also be transcribed in the recombinant strain, the fermentation and extraction of seven recombinant and

three control clones was undertaken to detect the presence of the diterpene terpentecin using three analytical approaches: a visual inspection of HPLC traces with MS detection, single ion monitoring for mass to charge ratio consistent with that of terpentecin, and finally a metabolomic analysis of the mass spectral data. The native terpentecin producer, *K. griseola* MF730-N6, was also included.

Following a 7-day culture in both ISP2 and lean medium at 30°C and 200 rpm, the fermentations were extracted with EtOAc and the crude extracts were analysed by LC-HRMS. A comparison of the ELSD chromatograms between seven recombinant strains carrying the terpentecin gene cluster and three control strains fermented in the same medium was performed to determine if major differences could be detected visually. Most ELSD traces of crude extracts from cultures grown in ISP2 showed an initial peak at 0.5 min, indicating the presence of salts in the extracts, followed by a single peak at 2.58 min, indicating that these extracts contained mostly one major compound.

Most ELSD traces of the lean medium looked very similar to one another; an initial salt peak at 0.5 min, followed by a chromatogram devoid of any significant peaks, indicating that most extracts did not contain any compounds in high abundance. An exception to these observations was the ELSD chromatograms of the extracts from recombinant strain 9 grown in the lean medium (Figure 2.27). These extracts contained two major compounds which did not appear in the extract of the control strains grown in the same medium. Unfortunately, the mass spectra for both recombinant and control strains were very similar at 2.60 and 2.88 min (Figure 2.28), and thus it was unclear which ions were contributing to the differences in ELSD intensities. It is possible that the compounds detected at those times did not ionize well and thus did not provide good mass spectral data. The same ELSD peaks were however present in one replicate of the control strain 1 grown in ISP2 medium (data not shown); it is possible that these compounds were not a product of the heterologous terpentecin genes, but rather they may be compounds that were variably produced by *S. lividans* TK24 ΔactΔred.

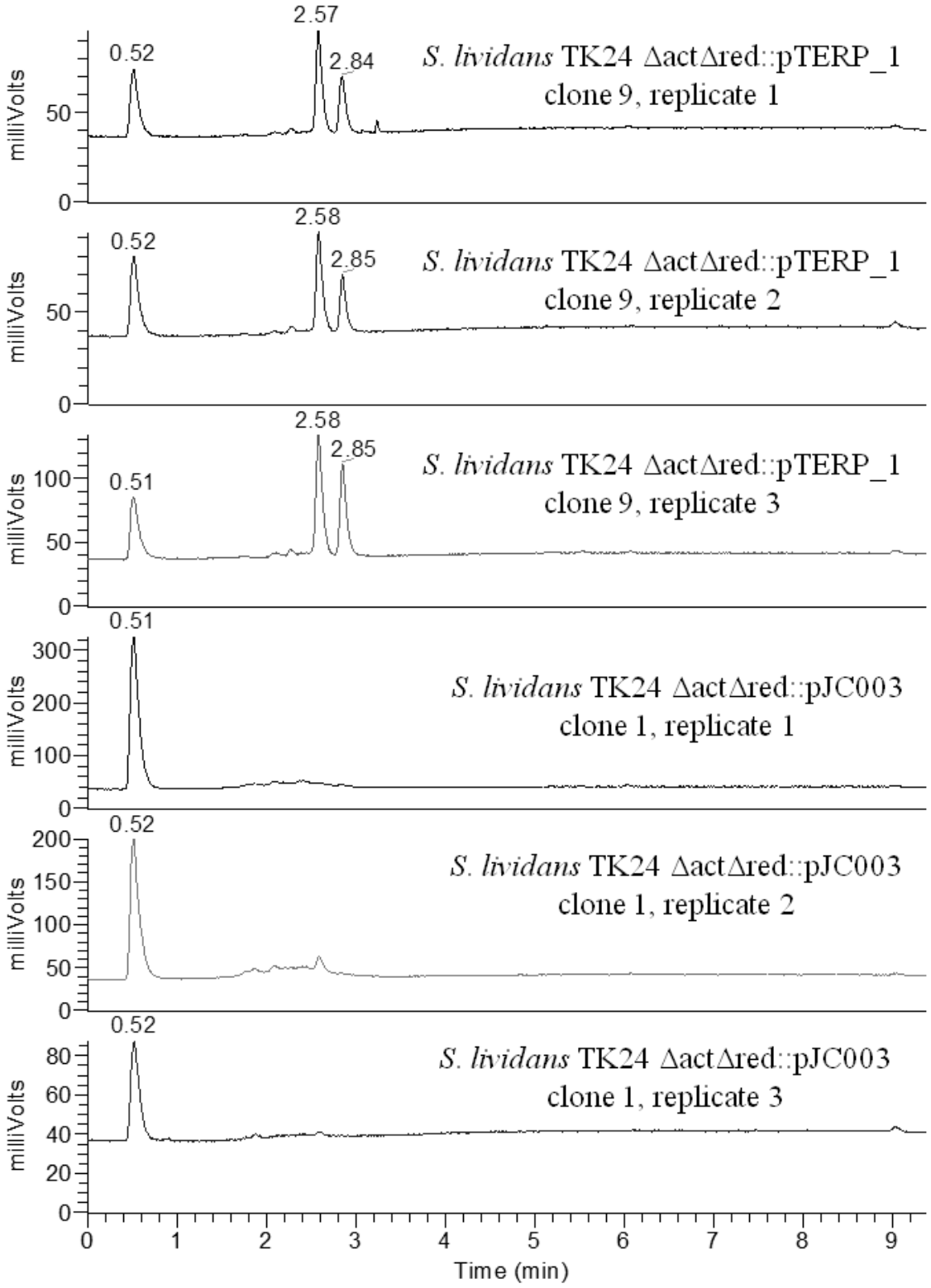


Figure 2.27. Evaporative light scattering detector (ELSD) chromatograms showing the most obvious differences between crude extracts from recombinant strain (clone 9) and crude extracts from control strain (clone 1) grown in lean medium.

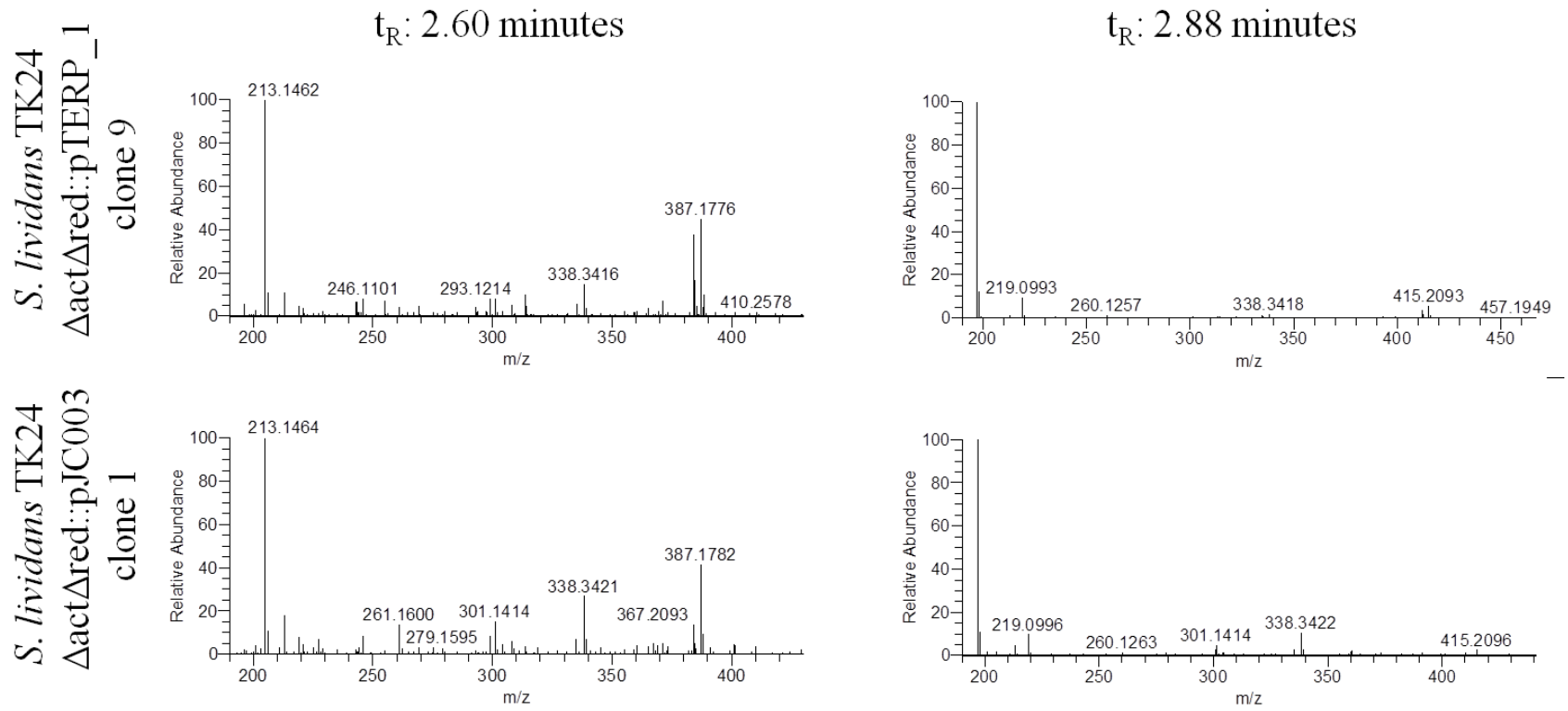


Figure 2.28. Selected portion of mass spectra for recombinant strain clone 9 and control stain clone 1 at 2.60 and 2.88 min illustrating that although the recombinant strain contained large evaporative light scattering detector peaks at these retention times, the mass spectra did not appear to contain any differing ions, suggesting that perhaps the compounds present in the recombinant strain did not ionize well, t_R: retention time, *m/z:* mass to charge ratio.

The second approach involved the monitoring of single ions consistent with that of the pseudomolecular $[M+H]^+$ of terpentecin (*m*/*z* of 365.1959 ±5 ppm) within the mass spectra to assess whether the recombinant strain could produce terpentecin. This analysis revealed that this mass to charge ratio was not helpful to predict the production of terpentecin since it was also found in the control strains containing the empty vector (Figure 2.29). In comparison to extracts from the native terpentecin producer, *K. griseola* MF730-N6, ions with *m*/*z* 365.1959 had slightly different retention times (t_R) ; the ions were detected at 2.85 min in *K. griseola*, while being detected around 2.61 min in the recombinant and control strains, respectively. The presence of this ion in recombinant and control strains most likely represents a compound with either the same molecular formula as terpentecin or a compound with a different molecular formula that gave rise to a molecular ion within 5 ppm of that from terpentecin. If another peak was observed in chromatograms of the recombinant strains at 2.85 min, a better chromatographic separation method could have been developed and in combination with tandem mass spectrometry experiments (MS^n) could have allowed the differentiation between the two compounds, but no such peak was observed. There was no evidence to suggest that this ion indicated the presence of terpentecin.

The third strategy to detect terpentecin in recombinant strains involved a metabolomic analysis to list all ions present in the extracts above the peak intensity threshold of 1E3. Although the analysis was performed on all seven recombinant clones grown in both media, only clone 9 grown in lean medium had mass spectral ions that were absent from the control strains. Seventeen ions, listed in Table 2.7, were present in triplicate extracts of the recombinant strain clone 9 grown in the lean medium that were also absent from all control strains of the same medium. The average retention time of these ions was 2.52 min and the *m*/*z* ranged between 203.1796 and 412.7023. Since a standard of terpentecin could not be obtained, the ions resulting from the fragmentation of the desired compound and their retention times are unknown and thus

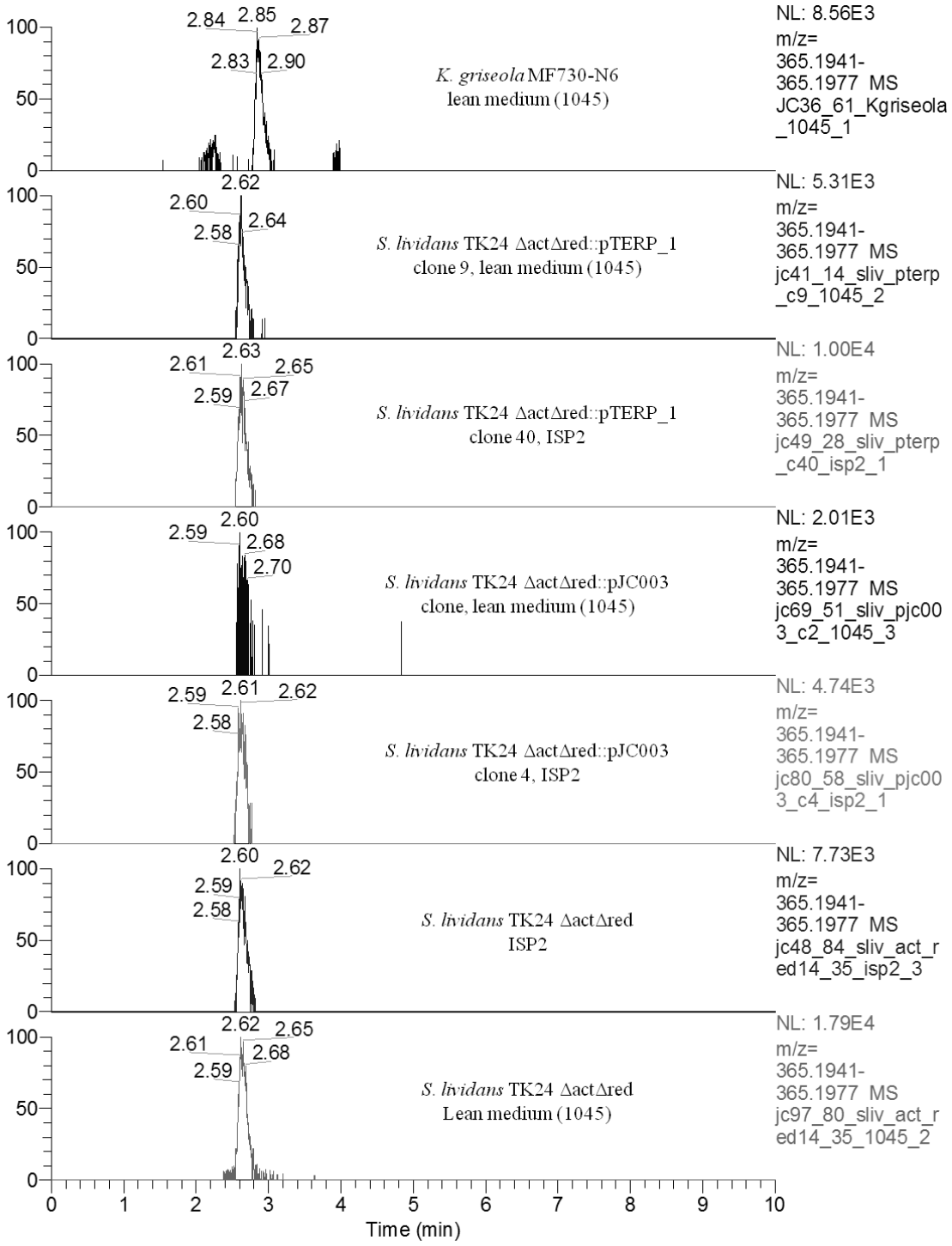


Figure 2.29. Chromatograms from single ion monitoring for ions 365.1959±5 ppm in extracts from the native terpentecin producer (*K. griseola*), recombinant strain (*S. lividans* TK24 ΔactΔred::pTERP_1), vector control strains (*S. lividans* TK24 ΔactΔred::pJC003) and the heterologous strain (*S. lividans* TK24 ΔactΔred); the terpentecin $[M+H]^+$ pseudomolecular ion mass to charge ratio was not an appropriate predictor of terpentecin production since it was found in the control strains.

Table 2.7. Ions present in extracts from *S. lividans* TK24 ΔactΔred::pTERP_1 clone 9 grown in lean medium that were absent from the control strains (*S. lividans* TK24 ΔactΔred::pJC003) grown in the same medium as determined by metabolomic analysis; *m*/*z:* mass to charge ratio, t_R: retention time.

m*/*z	**t_R (min)**
203.1796	3.08
251.1504	0.65
257.1286	2.75
263.1155	2.95
279.1104	2.74
279.1340	1.50
284.1969	1.76
311.1692	2.91
314.1448	2.98
325.1069	1.67
334.659	2.89
335.1605	2.86
335.2216	2.59
391.2093	2.85
405.1915	2.88
412.2007	2.88
412.7023	2.89

it was difficult to determine whether the ions listed here were fragments/derivatives of terpentecin. Knock-out experiments were thus undertaken in order to determine which ions in the native producer's extracts could be correlated with the production of the compound terpentecin.

2.3.4 Generation of two knock-out strains: *K. griseola* Δbaf and *K. griseola* Δterp

To abolish the production of bafilomycins and terpentecin in *K. griseola* MF730-N6, the removal of most of the bafilomycin biosynthetic genes and two essential terpentecin biosynthetic genes were carried out using the constructed deletion plasmids pBAFKO and pTERPKO. The bafilomycin knock-out strain was developed in order to assess the involvement of this known antibiotic in agar-based disk diffusion assay of *K. griseola* extracts against *Bacillus subtilis* and to identify which ELSD peaks and mass chromatogram ions belonged to the bafilomycins. Plasmids were constructed using the suicide plasmid pSET151 as the vector backbone along with the apramycin resistance gene flanked by left and right homologous arms. The suicide plasmid pSET151 was used as the backbone of the deletion plasmids due to its inability to replicate in a *Streptomyces* host and a thiostrepton resistance marker, making it suitable for rapid detection of single crossover mutants by resistance to thiostrepton in *Kitasatospora*[69, 78].

E. coli harbouring pBAFKO were first subjected to colony PCR and17 of the 20 clones screened contained a 3 kb amplicon (Figure 2.30). Plasmid DNA from clone number 12 was then subjected to restriction analysis with PstI/SacI and AatII. At the time of this experiment, the DNA sequence of the bafilomycin gene cluster in *K. griseola* M730-N6 was not known; preliminary *K. griseola* sequence data along with sequence data from *K. setae* had allowed the design of primers to amplify the left and right homologous arms but the exact location of restriction sites were unknown. The restriction analysis was thus ineffective at confirming proper assembly. However, a diagnostic 750 bp fragment resulting from SacI digestion (Figure 2.30), confirmed that the apramycin resistance gene was present

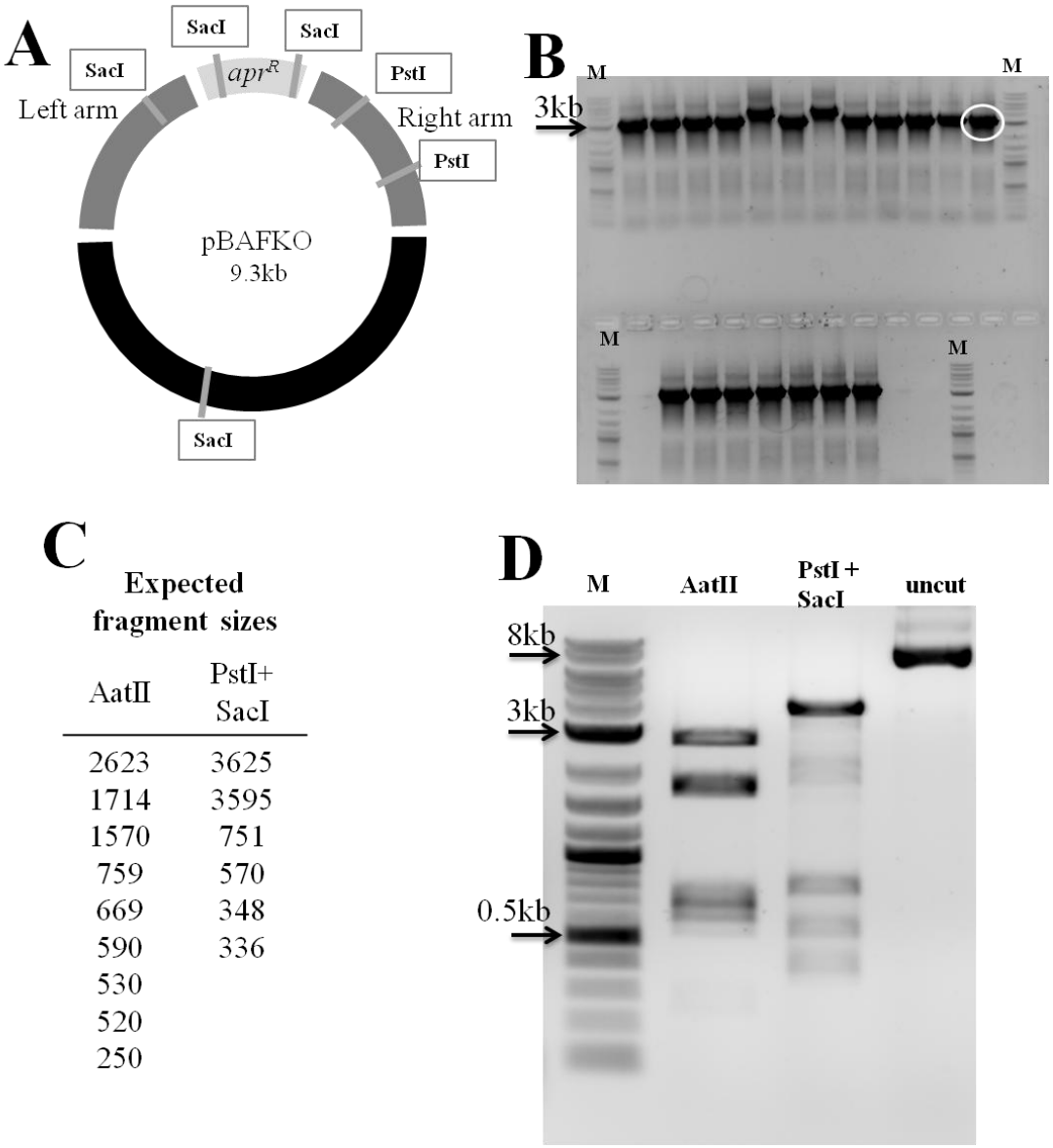


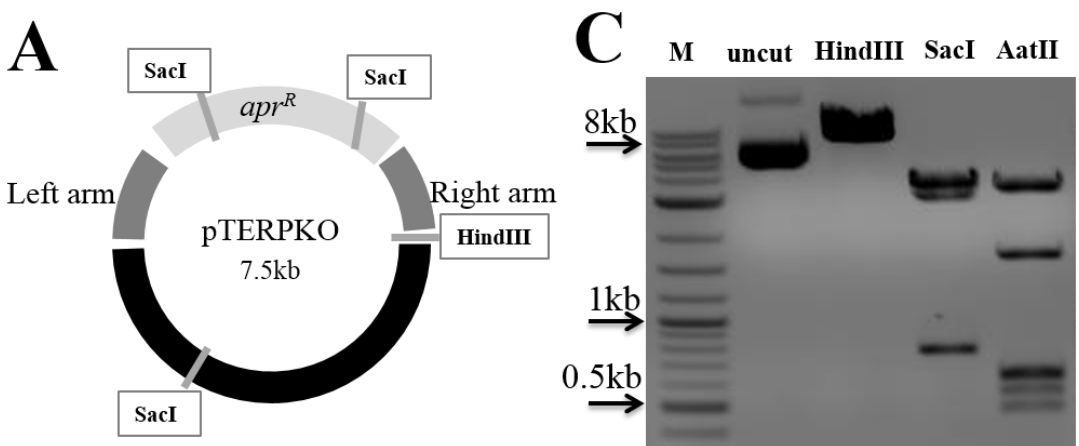
Figure 2.30. Restriction map of pBAFKO **(A)**, PCR screening of 20 *E. coli* pBAFKO transformants (clone 12 is circled) **(B)**, the expected fragment sizes for each restriction digest **(C)** agarose gel of pBAFKO clone 12 restriction digests **(D)**; M: 2-log ladder.

in pBAFKO clone 12. Subsequently, the bafilomycin biosynthetic gene cluster sequence was determined in *K. griseola* and it was confirmed that the fragment sizes obtained during the restriction digest were accurate. This clone was thus used in subsequent steps to generate the bafilomycin knock-out strain.

Similarly, pTERPKO was properly formed in clone 1 (Figure 2.31) as a fragment of approximately 8 kb was observed when linearized with HindIII, fragments of 3-4 kb and 700-800 bp when restricted with SacI and finally fragments of approximately 3 kb, 1500 bp, 600 bp, 550 bp and 500 bp following restriction by AatII. Since the observed bands were consistent with the predicted fragment sizes, this clone was used in subsequent steps to generate the terpentecin knock-out strain.

To identify baf and terp mutants, gDNA from putative mutant colonies were screened by PCR. Since a very large region was removed from the genome in *K. griseola* Δbaf, two different PCR reactions were necessary to confidently assign the knock-out genotype. The first PCR reaction used primers baf_Z_F and baf_Z_R which annealed within the cluster; a 759 bp amplicon was expected in wild-type strains and no amplicon was expected for knock-out strains since this region was absent. The gel picture (Figure 2.32 (A)), confirmed that the bafilomycin genes were still present in clone 6 as well as in the wild-type and suggested that they were absent from *K. griseola*::pBAFKO clones 1, 2, 3, 4, 5 and 7. The second PCR reaction used primers bafKO_left_F and bafKO_right_R which annealed at the outwards boundaries of the homologous arms; no amplicon was expected from the wild-type strain due to the large distance (73 kb) between the two primer annealing sites and a 3112 bp amplicon was expected in knock-out strains. The 3 kb amplicon (Figure 2.32 (A)) confirmed gene deletion in *K. griseola*::pBAFKO clones 3, 4, 5 and 7. *K. griseola*::pBAFKO clone 7 was renamed *K. griseola* Δbaf.

Similarly, PCR was used to determine the genotype of putative *K. griseola*::pTERPKO clones using primers SEQTERP-1-d5-F and SEQTERP-2-d5-R; a 3856 bp amplicon was



B

Expected fragment sizes

HindIII	SacI	AatII
7500	3521	3446
	3263	1570
	791	630
		590
		520
		250

Figure 2.31. Restriction map of pTERPKO **(A),** the expected fragment sizes for restriction digests **(B)** agarose gel of pTERPKO clone 1 restriction digests**(C)**; M: 2-log ladder.

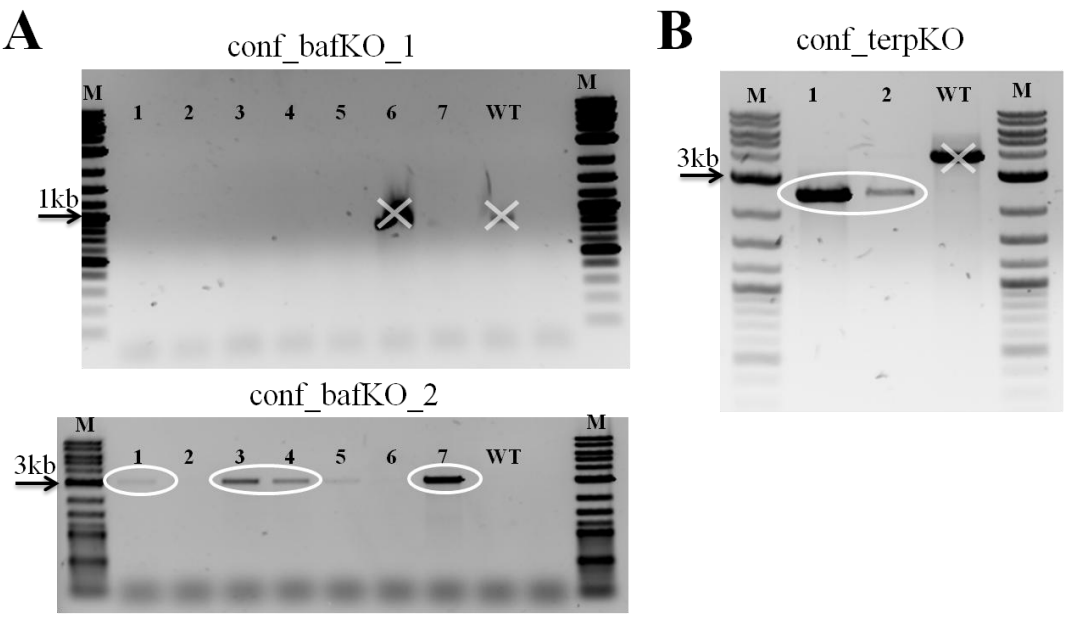


Figure 2.32. PCR identification of bafilomycin **(A)** and terpentecin **(B)** mutants; bands that correspond to a gene deletion are circled in white while those indicative of the wild-type genotype are marked with an X; the numbers refer to the clone number during the gene deletion process, while WT refers to the wild-type organism *K. griseola* MF730-N6 containing intact genes, M: 2-log ladder.

expected in the wild-type strain while a 2498 bp amplicon was expected in the knock-out strains. The genotype of *K. griseola*::pTERPKO colonies 1 and 2 (Figure 2.32 (B)) was consistent with the removal of *ORF11* and *ORF12* (partial) from the terpentecin biosynthetic gene cluster due to a 2.5 kb amplicon whereas amplification from wild-type genomic DNA resulted in a 4 kb amplicon. *K. griseola*::pTERPKO clone 1 was thus renamed *K. griseola* Δterp.

Sequencing of the 2500 bp amplicons from *K. griseola*::pTERPKO colonies 1 and 2 confirmed the replacement of *ORF11* and *ORF12* with the apramycin resistance gene. The sequenced amplicon from clone 1 matched the regions flanking the left and right homologous arms with an apramycin resistance gene at its centre; the sequence revealed that the apramycin resistance gene had been incorporated into pTERPKO in the forward direction. The same results were obtained for amplicons generated from *K. griseola*::pTERPKO clone 2. Although it is preferable to have two constructs of the deletion plasmid, one for each direction of transcription of the resistance gene, only the forward direction was obtained here and thus we cannot dismiss polar effects caused by this directionality on downstream genes. However, since the purpose of this experiment was not to confirm the role of *ORF11* and *ORF12* gene products in terpentecin biosynthesis but rather to cease the production of terpentecin in the organism, a polar effect here would not interfere with this goal. The same was true for the construct of the bafilomycin knock-out strain where the directionality of transcription of the apramycin resistance gene was not studied.

Assessing bafilomycin and terpentecin production in wild-type and knock-out strains

From the literature, *K. griseola* MF730-N6 was only known to produce two secondary metabolites: bafilomycin B1 (setamycin) and terpentecin[67, 86]. It is however clear from the ELSD of the HPLC trace (Figure 2.33 (A)) that there are far more than two secondary metabolites produced by this strain. In order to conclusively assign known compounds to ELSD peaks, two knock-out experiments were undertaken.

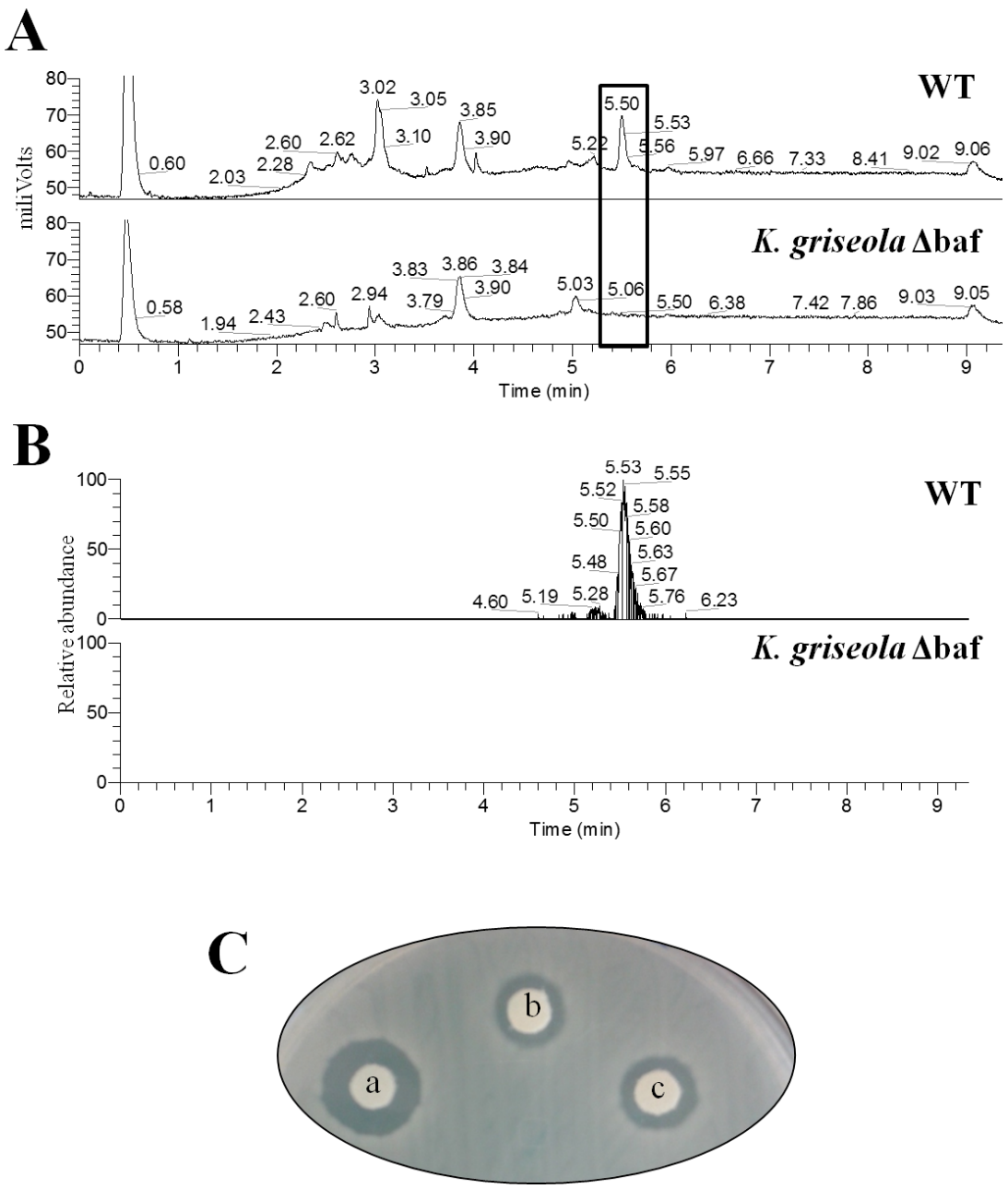


Figure 2.33. Evaporative light scattering detector (ELSD) chromatograms **(A)** and single ion monitoring of *m*/*z* of 838.4348 ± 5 ppm in HRMS chromatograms **(B)** for compounds produced in the wild-type *K. griseola* MF730-N6 and knock-out strain *K. griseola* Δbaf. The black box is placed on the region where metabolites disappeared in the knock-out strain, and inhibitory activity of 20 µL of fermentation broth from the wild-type strain (a) and knock-out strains (b,c) against *B. subtilis* **(C)**.

By removing most of the bafilomycin biosynthetic genes from the genome of *K. griseola* MF730-N6, a clear phenotypic difference was seen in the metabolome of the knock-out strain (Figure 2.33 (A)), where the ELSD peak present at 5.50 min in the wild-type is no longer present in fermentation extracts of the knock-out strain. Since HRMS data was also collected simultaneously, confirmation that the compounds detected at this time were in fact bafilomycin was obtained by scanning the mass spectra for a mass to charge ratio (*m*/z) of 838.4348 ± 5 ppm which represented the sodium adduct pseudomolecular ion $[M+Na]^+$ of bafilomycin B1 (Figure 2.33 (B)) where these ions were present in the wild-type strain and absent from the knock-out strain.

The inhibitory activity present in fermentations of *K. griseola* MF730-N6 was not caused solely by the presence of bafilomycins since growth inhibition in *B. subtilis* was still observed in the knock-out strains (Figure 2.33 (C)). The reduction in inhibition zone size may be caused by the slight inhibitory properties of bafilomycins against *B. subtilis* or by variations in yields of bioactive metabolites produced per 10 mL culture. This experiment confirmed the identity of the metabolites detected at 5.5 min in crude extracts of *K. griseola* as the bafilomycins and that they were only partly or not at all responsible for the inhibitory activity observed in crude extracts against *B. subtilis*.

When the essential terpentecin genes *ORF 11* and *ORF 12*, encoding for diterpene cyclases, were removed from the genome of *K. griseola* MF730-N6, it was unclear from the HPLC-ELSD chromatograms (Figure 2.34 (A)) which metabolites were no longer produced in the knock-out strain. A metabolomic analysis using a threshold value of 1E3 was thus undertaken. Rather than transforming the data to a binary data set as described in Forner et al.[24], the peak areas were kept for this study. This permitted a simpler visualization of which compounds were most affected by the knock-out of terpentecin genes.

From the Excel® data sheet generated from the metabolomic analysis, 53 ions (Table 2.8.) were determined to be present in all wild-type fermentation extracts and absent from all

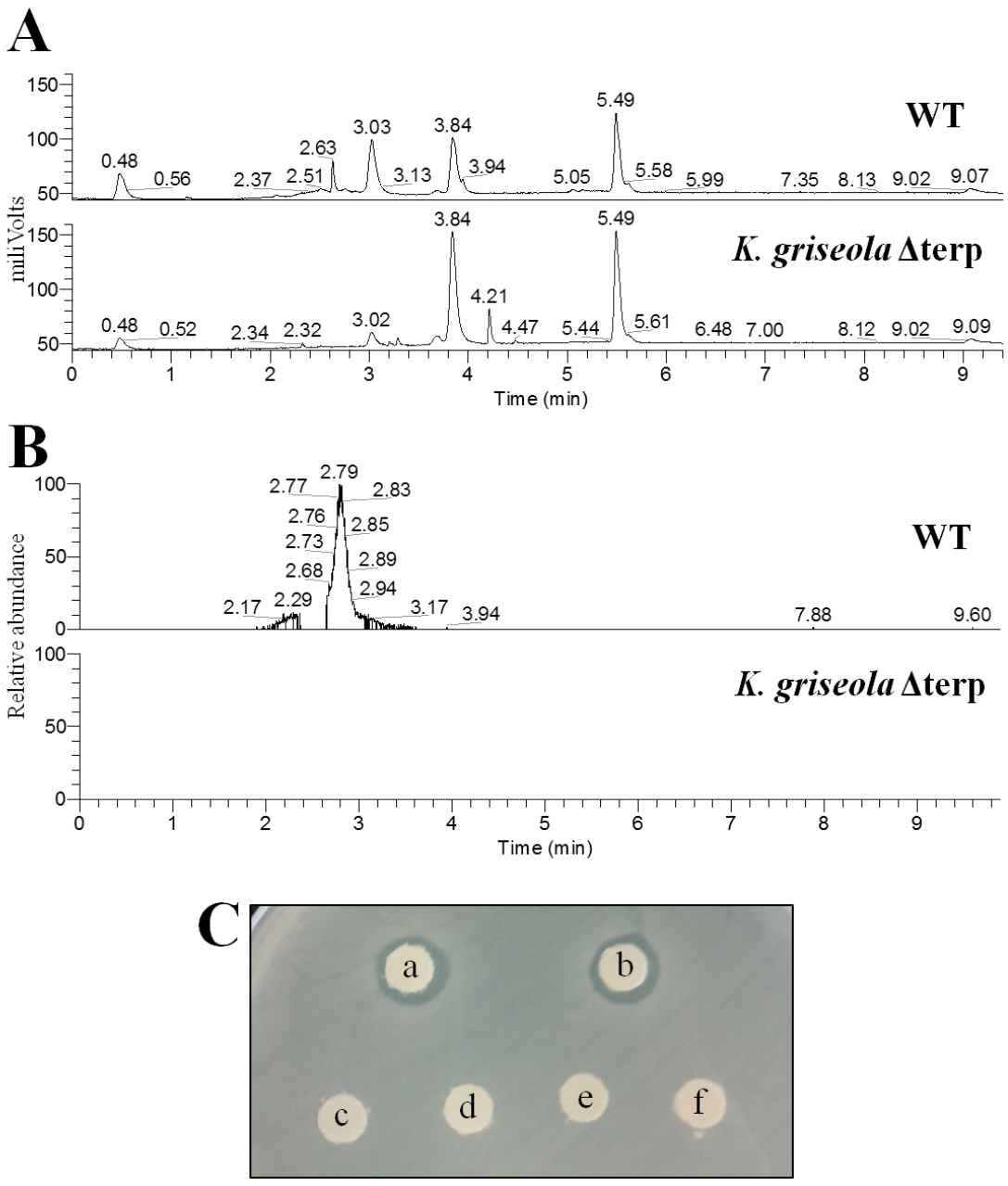


Figure 2.34. Evaporative light scattering detector (ELSD) chromatogram **(A)** and mass chromatogram selecting for ions of m/z of 365.1959 ± 5 ppm **(B)** for compounds produced in the wild-type *K. griseola* MF730-N6 and knock-out strain *K. griseola* Δterp and inhibitory activity of 20 µL of fermentation broth from the wild-type strain (a,b) and knock-out strains (c-f) against *B. subtilis* **(C)**.

Table 2.8. Ions absent from fermentations of *K. griseola* Δterp as determined by metabolomic analysis. WT: wild-type, *m/z:* mass to charge ratio, t_R: retention time, the $[M+H]^+$ pseudomolecular ions predicted for terpentecin[a], UCT-4B[b] and the *O*-methyl derivative of terpentecin[c] are noted.

Ions present in all 4 WT fermentations				Ions present in 3 WT fermentations	
m/z	**t_R (min)**	***m/z***	**t_R (min)**	***m/z***	**t_R (min)**
201.1278	3.08	419.2041	3.07	259.1147	3.08
205.1588	2.80	430.2436	2.52	263.1642	2.86
313.1797	3.08	432.2593	2.76	286.2126	2.86
329.1748	2.80	435.1992	2.55	289.1775	3.04
331.1904	2.64	437.2148	2.74	311.1641	2.76
333.2060	2.61	442.2803	3.08	315.1952	2.68
343.1902	3.07	457.2617	2.56	352.2481	3.43
345.1698	2.45	459.2777	2.39	381.227	2.58
347.1856	2.80	473.2568	2.40	384.2377	2.95
349.2010	2.63	475.2727	2.33	386.2524	2.67
351.2164	2.70	484.2906	3.07	389.1933	2.92
361.2010	3.08	485.2203	2.51	391.2094	2.67
363.1802	2.48	487.2362	2.38	407.2042	2.60
365.1957[a]	2.80	489.2521	2.35	411.2374	2.73
368.2429	2.81	491.2679	2.28	413.2172	2.28
373.1987	2.83	501.2515	2.48	421.1834	2.35
377.1958	2.53	503.2312	2.04	448.2542	2.25
379.2114	2.58	505.2472	2.35	477.2891	2.27
379.2482	3.10	513.2321	2.62	483.2409	2.91
381.1908[b]	2.34	517.2462	2.48	500.2863	2.60
383.2064	2.82	519.2623	2.41	502.3015	2.75
395.2063	2.52	521.2419	2.00	527.2484	2.57
397.2219[c]	2.77	531.2429	2.37	536.2891	2.92
400.2329	2.81	535.2572	2.08	541.2444	2.93
405.1885	2.81	545.2592	2.57	544.2637	2.60
414.2496	3.08	557.2382	2.54	552.2841	2.45
416.2279	2.37			559.2747	2.62
				561.2535	2.14

those of terpentecin knock-out strains; all were detected between 2.00 and 3.43 min. From the ELSD chromatogram (Figure 2.34 (A)), there was a very complex mixture of compounds detected between 2 and 3.5 min. This complex mixture may be a consequence of the dimer and hemiacetal tautomers formed when terpentecin was analysed previously[87, 88]. This may also explain why previous attempts to isolate terpentecin from the native hosts were unsuccessful using our chromatographic separation and detection methods (data not shown).

Clerocidin, a clerodane diterpene very similar to terpentecin, differing only in the substituents on the decalin ring, was previously reported to be easily transformed from its dimeric form into the *O*-methyl derivative in methanol[87]. Given its structural similarity to terpentecin and since the crude extracts were also dissolved in methanol for LC-HRMS analysis, it is possible that terpentecin had been transformed into its *O*-methyl derivative (Figure 2.35). In fact, ions consistent with both the pseudomolecular ion $[M+H]^+$ and the sodium adduct $[M+Na]^+$of the *O*-methyl derivative were highlighted (Table 2.8) as ions present in wild-type extracts and absent from all knock-out extracts (calculated *m*/*z*: 397.2202 and 419.2040, respectively) during the metabolomic analysis.

A further 28 ions were added to the list by using ions which were found in three of the four wild-type fermentations while still being absent from all eight knock-out fermentations. Present in these lists were ions with *m*/*z* consistent with the pseudomolecular ion $[M+H]^+$ of terpentecin (365.1957) and *m*/*z* 381.1908, which was consistent with that of UCT-4B, a compound almost identical to terpentecin, differing only in the presence of an extra hydroxyl group (Figure 2.35). Since HRMS data rarely provide structural information past the molecular formula, it remains unclear whether *K. griseola* is capable of producing both terpentecin and UCT-4B or whether it is producing another hydroxylated derivative of terpentecin; NMR experiments on the purified compound would provide this information. To date only *Streptomyces* sp. s464 has been reported to produce both terpentecin and UCT-4B[89]. Only two more suggestions regarding the listed ions' possible molecular compositions could be

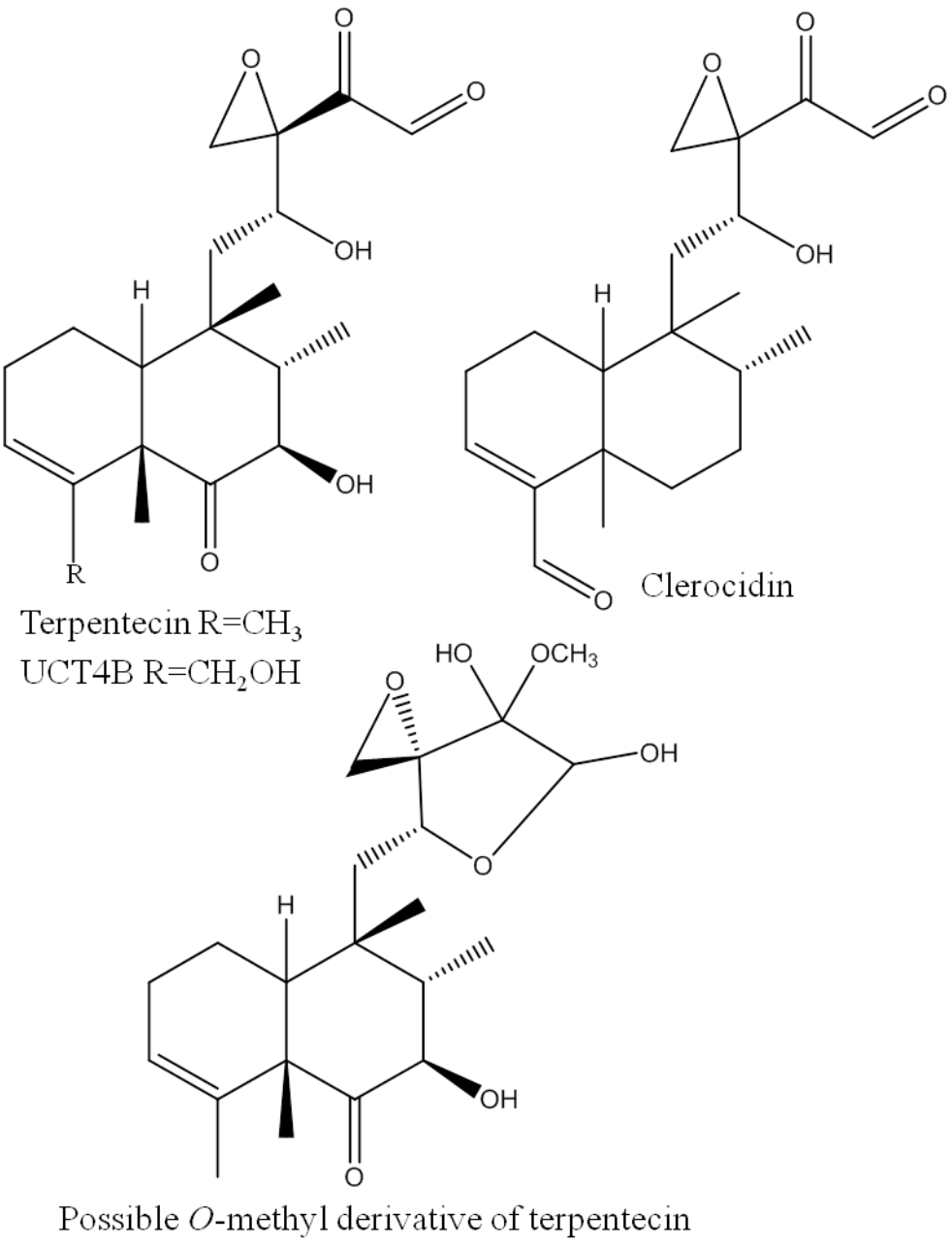


Figure 2.35. Structures of terpentecin and related structures UCT4B and clerocidin as well as its possible *O*-methyl derivative when dissolved in methanol.

made: ions with *m*/*z* 379.2114 and 395.2063 may be analogues of terpentecin and UCT-4B where an extra methyl group would be present somewhere on the molecules (calculated $[M+H]^+$ 375.2115 and 395.2064, respectively).

A principal component analysis of the HRMS ions present within each sample was used to rapidly identify which ions contributed most to the variability between wild-type and knock-out crude extracts. The analysis revealed that the chemical composition of the terpentecin knock-out strains was clearly different from that of the wild-type *K. griseola* strain as seen in the scores plot of principal components 1 and 2 (Figure 2.36 (A)). The ions responsible for this differentiation are shown in the loadings plots of the same principal components (Figure 2.36 (B)) and included the following *m*/*z*: 487.2362, 519.2623, 505.2472, 503.2312, 535.2572, 489.2521, 379.2144, 365.1957, 414.2496 and 419.2041, all of which had been previously highlighted from the Excel® data sheet analysis.

From the Excel® and principal component analyses, it was now possible to restrict the range of *m*/*z* in the mass chromatograms (Figure 2.34 (B)) where ions consistent with the $[M+H]^+$ pseudomolecular ion of terpentecin were detected in the wild-type and were absent in the knock-out strains.

The lack of production of terpentecin in the knock-out strains was further supported by the lack of growth inhibition of *Bacillus subtilis* by 20 µL of broth and crude extract (0.5 mg/mL) of knock-out strains. Although it was known that *K. griseola* produced another antibacterial compound, bafilomycin B1, it was clear that the majority of the *B. subtilis* inhibitory activity in the broth and crude extracts of *K. griseola* MF730-N6 was due to the presence of terpentecin since the knock-out of essential terpentecin genes completely abolished any inhibitory activity against *B. subtilis* (Figure 2.34 (C)).

This observation was consistent with the reported minimal inhibitory concentrations (MIC) of <0.05 µg/mL and 25 µg/mL for terpentecin and bafilomycin B1 respectively, against *B. subtilis* PCI 219[67, 90] and by previous experiments where the same lack of growth inhibition

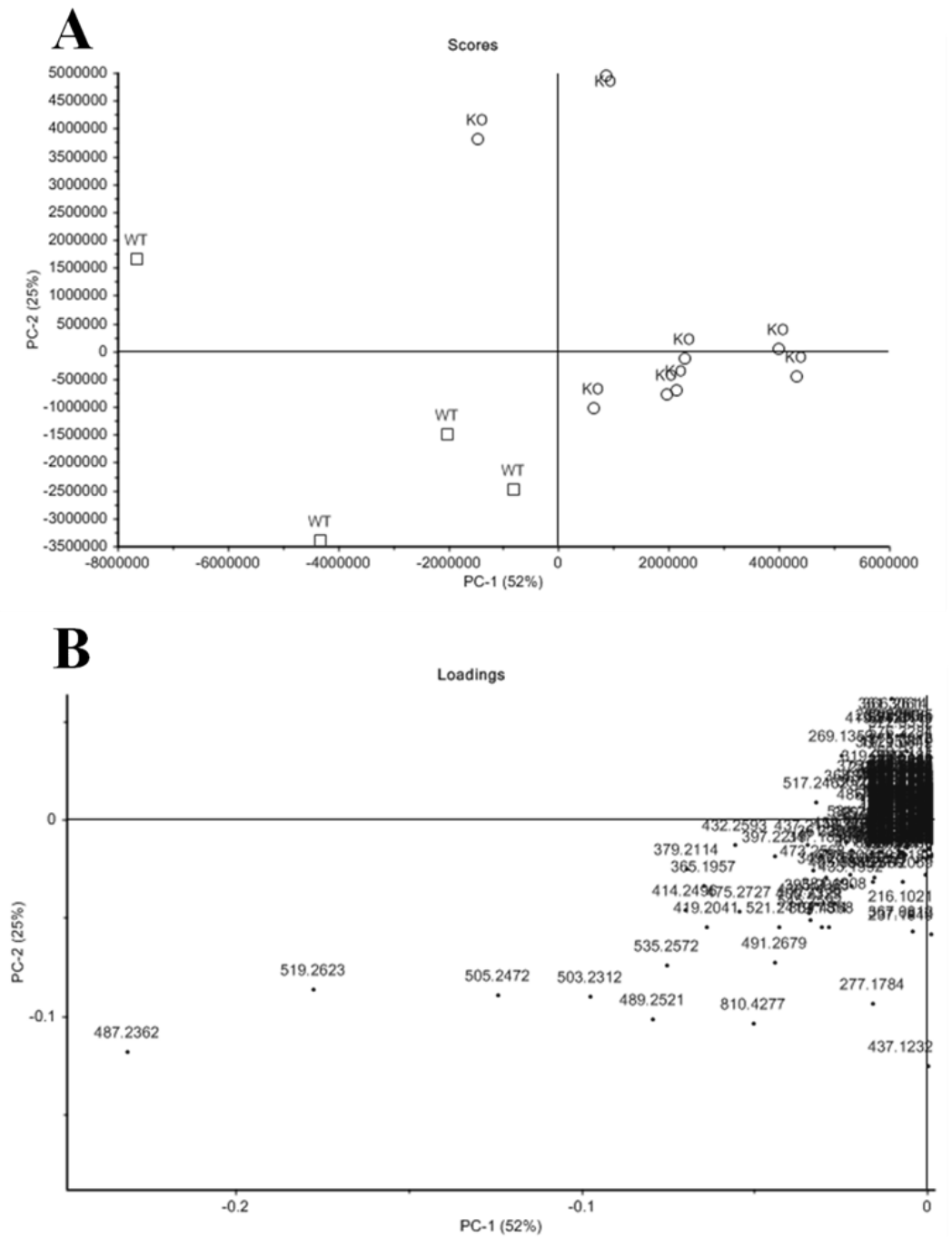


Figure 2.36. Scores plot **(A)** and loadings plot **(B)** of the principal component analysis highlighting the differences in ions present in the wild-type (WT) and terpentecin knock-out (KO) strains.

towards *B. subtilis* was observed when the same genes were inactivated in *K. griseola* MF730-N6[64]. The reduced bioactivity in the bafilomycin knock-out strain may be a reflection of the yields of terpentecin in the fermentation broths since the latter is at least 500 times more potent than bafilomycins. It is however also possible that there were other bioactive compounds that were not consistently being produced and were thus confounding the results or that there is synergism between terpentecin and bafilomycins. This can only be confirmed by fully characterizing the metabolome and the genome of this strain.

Although a complementation study was not performed here, there is evidence to suggest that the knock-out phenotype was not a result of polar effects causing failed transcription in genes at other loci on the genome. Firstly, there were no apparent differences in growth rates or sporulation between knock-out and wild-type strains and secondly, the knock-out strain was still capable of producing other secondary metabolites as evidenced by the peaks at 5.49 min which were proven to be bafilomycins in the *K. griseola* Δbaf strain.

Now that we were able to correlate the presence of ion *m*/*z* 391.2094 (t_R 2.67 min) with the ability of a strain to produce terpentecin, it may be suggested that the ion *m*/*z* 391.2093 (t_R 2.85 min) detected in recombinant fermentations (clone 9, lean medium) indicated heterologous terpentecin production. However, discrepancies between the retention times suggested that these two ions represented different compounds. The evidence was thus not sufficient to conclude that terpentecin was produced heterologously as only one terpentecin-relation ion was detected in only one replicate of one recombinant clone and the retention times did not match. Another fermentation study was thus carried out.

2.2.5 Reassessing terpentecin production in *S. lividans* TK24 ΔactΔred::pTERP_1.

Now that the presence of various ions could be correlated with the ability of the organism to produce the diterpene terpentecin, another fermentation study was undertaken in order to find any terpentecin related ions in three recombinant strains *S. lividans* TK24 ΔactΔred::pTERP_1

using a greater variety of media: MYM, TSB, YEME+34% sucrose+0.5% glycine, lean medium, ISP3 and mannitol-soy.

The results were analysed individually by medium since the principal component analysis (Figure 2.37) revealed that samples (recombinant strain and control) were separating based on the fermentation medium used rather than by the genotype of the strain.

The Excel® data sheet was used to highlight which ions were present in extracts of the recombinant (*S. lividans* TK24 ΔactΔred::pTERP_1) and absent from that of the control strain (*S. lividans* TK24 ΔactΔred::pJC003) and media blank. Firstly, there were only 5 ions that were present in all three recombinant strain fermentations and absent from all control and media fermentation extracts: ions with *m*/*z* 257.1287 and 418.2799 from the lean medium and 577.5195, 586.5408 and 617.5120 from the fermentation medium mannitol-soy. An example of the mass chromatograms for the ion with *m*/*z* 257.1287 (Figure 2.38), where the presence of a new compound is detected in extracts from the recombinant strain while being absent from that of control extracts. Although none of the five new ions had been highlighted as being related to terpentecin in the previously mentioned knock-out experiment, ion with *m*/*z* 257.1287 had been previously highlighted during the first fermentation as a heterologous metabolite as it was present in all replicates of recombinant strain clone 9 extracts while being absent from the control strains (see Section 2.3.3).

It is possible that the compounds obtained from the recombinant strain are not identical to those produced by the native host. This has been observed previously in this strain when the diterpene cyclase genes of the terpentecin biosynthetic gene cluster were heterologously expressed in *S. lividans* along with GGPP synthase and the production of a new compound, terpentetriene, was observed[64]. It is thus possible that some of the highlighted ions may be analogues of terpentecin, intermediate compounds that were not biosynthesized by the entire terpentecin biosynthetic gene cluster or perhaps shunt metabolites, where biosynthetic intermediates were further modified into new compounds. It is also possible that the terpentecin

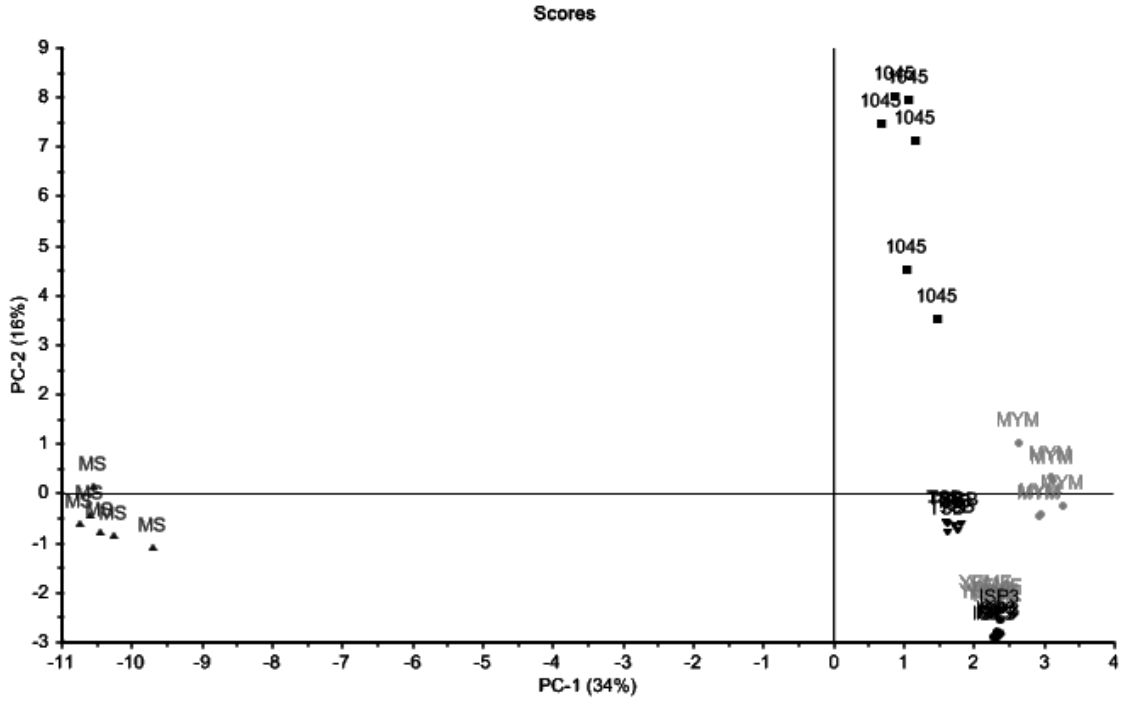


Figure 2.37. Scores plot of the principal component analysis using the entire data set to illustrate that the metabolites present in the crude extracts of the recombinant and wild-type strains differed mostly by fermentation medium used rather than by genotype.

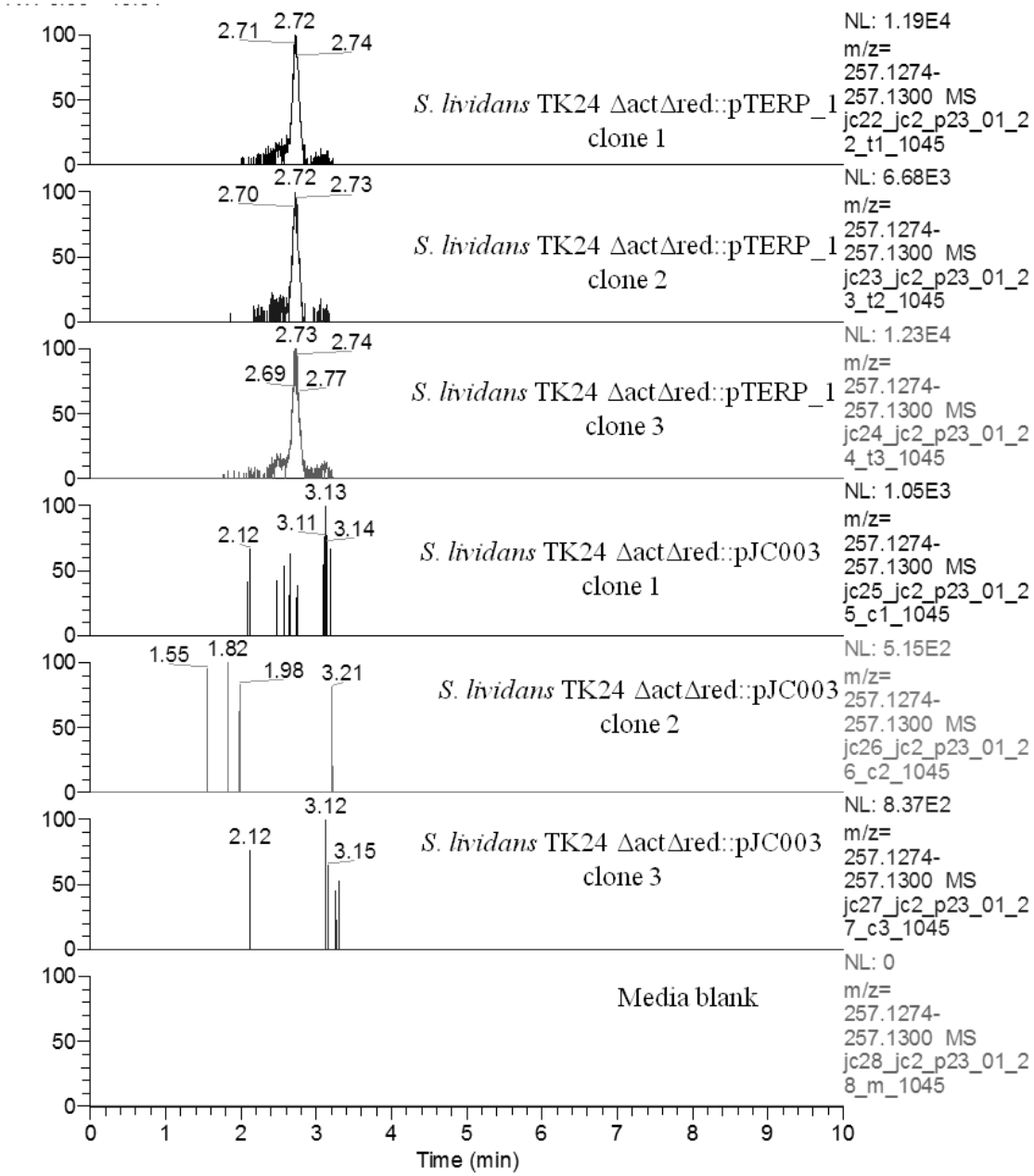


Figure 2.38. Screening for presence of heterologously expressed metabolites by single ion monitoring of *m/z* of 257.1286 ± 5 ppm in fermentation extracts of three colonies of the recombinant strain (*S. lividans* TK24 ΔactΔred::pTERP_1), three colonies of the control strain (*S. lividans* TK24 ΔactΔred::pJC003) and one media blank.

biosynthetic gene cluster was incomplete and that there were genes missing from the recombinant strain's genome that are resulting in partial biosynthesis. Although there are only two cytochrome P450 genes in the reported gene cluster, 6 oxygenation reactions are predicted; it is not possible to predict whether other P450 enzymes are necessary or whether the same enzyme can catalyze multiple oxygenations as is the case for PikC which carries out hydroxylations at two different sites of the pikromycin macrolactone core[91]. Since regions flanking the biosynthetic gene cluster did not seem to be involved in terpentecin biosynthesis (data not shown), it is also possible that other P450 enzymes in *K. griseola* were acting *in trans* to perform the additional oxygenation reactions. Knock-out experiments of each individual P450 gene within the terpentecin biosynthetic gene cluster would indicate the biosynthetic step(s) accomplished by each enzyme and until terpentecin is produced in a recombinant strain, the minimal biosynthetic gene cluster remains unknown.

The new compounds whose ions were detected in the extracts of recombinant strain fermentations would need to be purified and their structure elucidated in order to confirm structural similarity to terpentecin. Priority should be given to the compounds originating from the lean medium extract since those ions were not detected in any other medium, whereas those from mannitol-soy medium were also detected in control fermentations in ISP3 medium, making them less likely to represent compounds produced via the heterologously expressed genes; they were perhaps present below the detection level in the control strains cultured in mannitol-soy medium.

2.4 CONCLUSION

In attempting to develop a general host for the heterologous production of bioactive terpenes in high yield, numerous molecular tools were generated including the new integrating cloning vector pJC003 capable of cloning large biosynthetic gene clusters into *Streptomyces*, a new version of a known constitutive promoter to drive the transcription of downstream genes

and finally another version of *S. lividans* with a simplified metabolic profile for easy detection of heterologously expressed metabolites. Given that these tools were not commercially available, their development during this study has made them available to other members in our laboratory and collaborators for future genetic studies involving microorganisms of the genera *Streptomyces* and *Kitasatospora*.

In attempting to heterologously produce the diterpene terpentecin, great effort was dedicated to confidently detect terpentecin in fermentation extracts. Following the knock-out of essential genes in the native producer *K. griseola*, we were able to confirm that terpentecin was produced in this strain under our laboratory conditions and that LC-HRMS could be used to detect terpentecin, though in very low yields, via the monitoring of 106 ions, some with *m*/*z* consistent with the pseudomolecular ions $[M+H]^+$ for terpentecin, UCT-4B and the *O*-methyl derivative of terpentecin: 365.1959, 381.1908 and 397.2220, respectively.

Although the terpentecin biosynthetic gene cluster was successfully integrated in the genome of *S. lividans* TK24 ΔactΔred, there was insufficient evidence to confidently support heterologous production of terpentecin. In the first fermentation study, numerous colonies were cultured in two media and only clone 9 grown in the lean medium produced any metabolite absent from control fermentations. The second fermentation thus investigated fewer recombinant strains and instead increased the number of media used to six; two media showed metabolites not produced in control strains, lean medium and mannitol-soy. Lean medium thus seemed to be promising for the heterologous production of terpentecin as it is inexpensive, easy to prepare and provides a clean HRMS background; it was also the medium reported to culture *K. griseola* MF730-N6 for terpentecin production[67]. This medium could thus be used in future heterologous studies using *S. lividans* TK24 ΔactΔred for the heterologous expression of terpentecin or other target compounds.

Although the evidence presented here did not confirm the production of terpentecin in the recombinant strain, the in-depth analysis of numerous fermentations conducted during this study

has lead to the discovery of new natural products from members of the genus *Kitasatospora*; this will be the subject of the following chapters of this thesis.

CHAPTER 3 - ISOLATION AND STRUCTURE ELUCIDATION OF SATOSPORIN A AND B: NEW POLYKETIDES FROM THE ACTINOMYCETE *KITASATOSPORA GRISEOLA*

Collaborators and their contributions to this chapter:
Computational calculations (Sections 3.2.4 and 3.3.3) performed by Dr Jason Pearson, UPEI
Modified Mosher and Tanaka derivatizations performed by Dr Fabrice Berrué, UPEI
Antimicrobial assays performed by Martin Lanteigne, UPEI
Assay against healthy cell lines (HEKA and BJ) were performed by Martin Lanteigne, UPEI

This chapter is a modification of the material published as

Arens, J.C., F. Berrue, J.K. Pearson, and R.G. Kerr, Isolation and structure elucidation of satosporin A and B: new polyketides from *Kitasatospora griseola*. *Org. Lett.*, 2013. **15**(15): 3864-7.

3.1 INTRODUCTION

Actinomycetes are filamentous, spore-forming, Gram-positive bacteria of high G+C content associated with sediments of both terrestrial and marine sources[71]. They are the most prolific source of antibiotics and are responsible for the production of 45% of all reported bioactive metabolites of microbial origin[20, 92, 93]. Due to this prolific secondary metabolite production, there is much precedence for searching unexplored habitats for the discovery of novel actinomycetes[28].

Although this method has experienced success, there is also genetic evidence to support the observation that previously studied organisms have the potential to produce far more secondary metabolites than what has been reported. *Streptomyces coelicolor*, for example, has been well studied since the 1960s and had been reported to produce 4 antibiotics: actinorhodin, prodiginines, calcium-dependant antibiotic and methylenomycin[71]. It was not until its genome was sequenced by Bentley et al. in 2002 that its true metabolite production potential was revealed, indicating a total of 20 biosynthetic genes clusters[30]. Since then, the new peptide natural product coelichelin was successfully isolated from the organism using an iron-deficient culture medium[32].

The genus *Kitasatospora*, formely *Kitasatosporia*, contains 20 accepted species isolated mostly from soil samples from Japan, Korea, China, India and Brazil. The main characteristics separating members of this genus to other Actinomycetales are: the presence of both $_{LL}$- and *meso*-diaminopimelic acid (DAP) in the peptidoglycan layer of the cell wall at certain stages of differentiation, the presence of galactose in cell hydrolysates and their ability to form submerged spores in liquid cultures[86]. The genus was, however, so closely related to *Streptomyces* that a decade after its first description by Omura et al. in 1982[94], the two genera were unified based on partial 16S rRNA sequences being incapable of discriminating between the two[95]. This unification was short-lived as the genus *Kitasatospora* was re-established by Zhang et al. in 1997 when full length 16S rRNA and 16S-23S rRNA gene spacer sequences were used to

resolve the two genera[96]. Since then, evidence continues to support the differentiation of the two genera; for example, drastic differences can be observed in the amino acid sequences of SsgB, a protein involved in sporulation and cell division, between *Streptomyces* and *Kitasatospora* organisms[97].

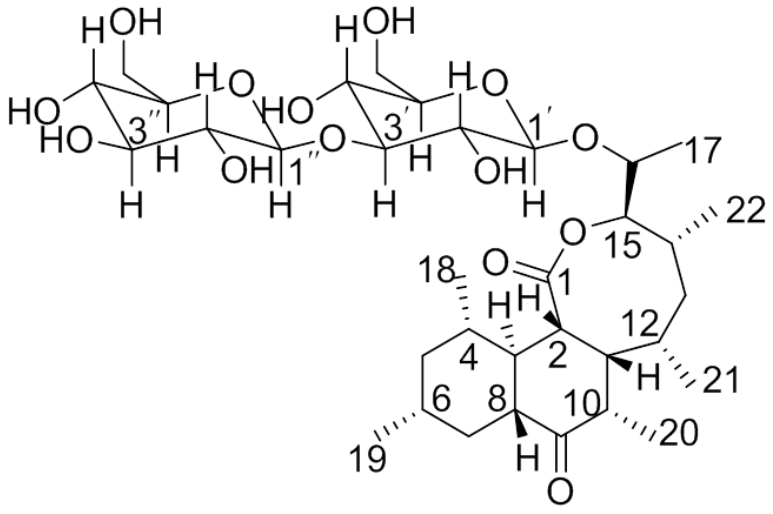
The first compound isolated from *Kitasatospora* was setamycin in the early 1980s[94]. Since then, numerous other compounds have been isolated from both accepted and proposed species including a bafilomycin-like compound, phosalacine, propioxatins, terpentecin, compound FR-900494, cystargin, tyrostatin, compound SUAM-20007-20012, phospholipase D, compound F-0368, kimorexins, fuzanins E-I and most recently endophenazine derivatives from *Kitasatospora* sp. HKI 714 in 2014[98-100]. The numerous compounds produced by this genus favour the possibility of discovering new natural products from fermentations of these bacteria.

In our efforts to identify new natural products, members of the genus *Kitasatospora* were cultured and this has led to the isolation of the novel polyketides, satosporin A (**1**) and B (**2**) (Figure 3.1). The aim of this research was to fully elucidate the structure of these compounds as well as the corresponding aglycone, satosporin C (**3**), and to briefly investigate their biological activity.

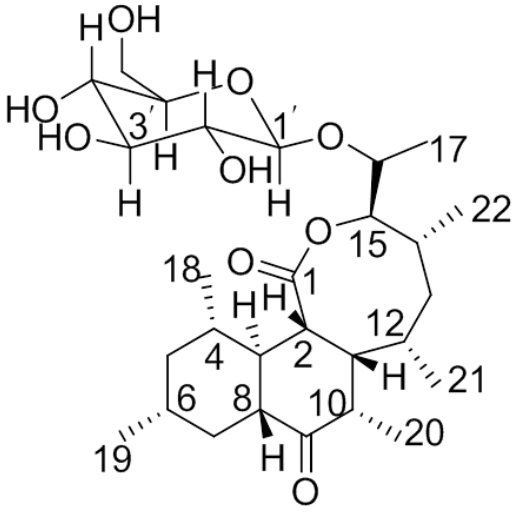
3.2 MATERIALS AND METHODS

3.2.1 Fermentation of *Kitasatospora griseola* MF730-N6

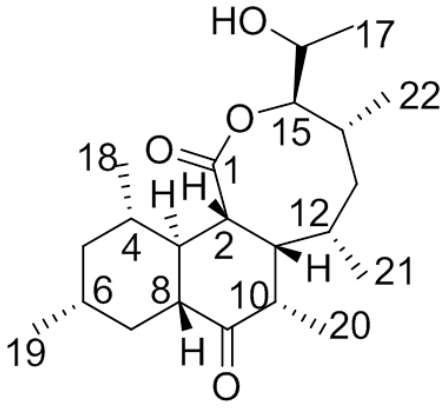
Seed cultures of *K. griseola* MF730-N6 were obtained first in 25 X 150 mm glass culture tubes and second in 250 mL Erlenmeyer flasks containing glass beads and 10 mL and 50 mL of sterile seed media containing 10 g of glucose and 10 g of yeast extract per liter. The first seed was inoculated with a scraping from an agar plate containing the sporulating organism, while the second seed was inoculated with 2.5 mL of seed 1. Seed cultures were allowed to grow for 24 h at 30°C in an orbital shaker at 200 rpm. A total of 24 fermentations were carried out in Fernbach flasks containing 500 mL of the lean production medium[67] containing 0.4 g glucose, 0.8 g galactose, 0.8 g maltose, 1.6 g dextrin, 0.8 g Bacto™-Soytone and 0.3 g $(NH_4)_2SO_4$ per liter



satosporin A (1)



satosporin B (2)



satosporin C (3)

Figure 3.1. Structures of satosporins A-C.

inoculated with 5 mL from the second seed and were grown for 48 h, 200 rpm and at 27°C to give a total volume of 12 L. A sample was taken from each fermentation and plated on agar to ensure a single bacterial colony type and thus no contamination of the cultures.

3.2.2 Isolation of satosporins from *K. griseola* fermentation

Isolation of satosporin A and B

Fermentations were extracted by shaking each 500 mL Fernbach for 1 h with 100 g activated Diaion® HP-20 resin which was then rinsed with 250 mL diH_2O and extracted using 250 mL 50% aqueous MeOH and finally 250 mL 100% MeOH. Extracts were then combined and solvents were removed using a rotary evaporator. The 100% MeOH extract was then sequentially partitioned using water and EtOAc followed by hexane and ACN. The ACN extract was then subjected to two consecutive orthogonal silica based fractionations. Firstly, the sample was loaded onto a 500 mg capacity Diol HyperSep™ SPE cartridge (Thermo Scientific) and fractions were obtained by eluting with 15 mL 80:20 hexane:*tert*-butyl methyl ether (TBME), 100% TBME, EtOAc and finally MeOH. Fraction 4 (MeOH) was then loaded onto a 500 mg capacity C_{18} HyperSep™ cartridge and fractions were obtained by eluting with 15 mL 95:5 H_2O: ACN, 70:30 H_2O: ACN, 50:50 H_2O: ACN, 100% ACN, isopropyl alcohol (IPA), acetone and finally 1:1 dichloromethane (DCM):MeOH. HPLC was then used to obtain pure compounds using a Gemini® 5 µm C_{18} column and a 30:70 H_2O:MeOH isocratic solvent system for 20 min followed by washing the column with 100% MeOH for 6 min and equilibrating the column back to the starting conditions for 9 min using 100 µL injections of a 10 mg/mL solution. Samples were subjected to LC-HRMS analysis at various steps along the purification process using Thermo Scientific™ LTQ Exactive equipment with a Core Shell Kinetex 1.7 µm C_{18}, 50 X 2.1 mm, 100 Å column, photodiode array detector, SEDEX 80 Sedere LT-ELSD evaporative light scattering detector and a Finnigan LXQ ion trap mass spectrometer in positive mode equipped with a electrospray ionization source detecting *m*/*z* between 190 and 2000. Chromatographic separation was achieved with a 4.8 min gradient of 95:5 water:ACN to 100% ACN, 3.2 min of

100% ACN and finally a return to starting conditions over 3 min; solvents contained 0.1% formic acid and a flow rate of 500 µL/min was used.

Satosporin A (**1**): colourless oil; $[\alpha]^{25}_{D}$: -5.2 (*c* 0.1, MeOH); IR ν_{max} 3363, 2952, 2919, 2875, 1736, 1704, 1592. HRESIMS *m/z* 711.3558 $[M+Na]^+$,(calculated for $C_{34}H_{56}O_{14}Na$ 711.3562, Δ = -0.6 ppm). MS^n: *m/z* 667.3663, 549.3041, 505.3146, 387.2524, 369.2419, 365.1072, 347.0968. NMR data are shown in Table 3.2.

Satosporin B (**2**): colourless oil; $[\alpha]^{25}_{D}$: -16.9 (*c* 0.04, MeOH); IR ν_{max} 3367, 2927, 1736. HRESIMS *m/z* 549.3021 $[M+Na]^+$, (calculated for $C_{28}H_{46}O_9Na$ 549.3040, Δ = -2.3 ppm). NMR data are shown in Table 3.3.

Obtaining the aglycone satosporin C by enzymatic deglucosylation

An enzymatic deglucosylation was attempted following the method of Shafiee et al.[101] using first endo-1,3(4)-β-glucanase (Sigma-Aldrich, G4423-100g) and then β-glucosidase from almond (Sigma-Aldrich, 49290-250mg). 10 µL of a 50 mg/mL solution of satosporin A in DMSO was incubated at 37°C in 990 µL 0.1 M sodium acetate buffer (pH 6.0) along with 10 mg enzyme in a microcentrifuge tube. 25 µL aliquots were taken at times 0, 0.5, 1, 2, 4, 8, 14, 24, 48, 96 and 192 h, added to 225 µL MeOH and kept at 4°C until the experiment was terminated. A sample was also taken prior to the addition of the enzyme in order to serve as a negative control.

The samples were analysed by LC-HRMS and the relative amounts of each analogue was assessed by creating a processing method in the Xcalibur™ software (Thermo Scientific™) to measure the area under the peak for the ions with *m/z* values of 711.3562, 549.3040 and 387.2511 (± 5 ppm) representing the $[M+Na]^+$ adducts for each analogue and using the sum of all three areas to represent each analogue as a percentage of total satosporins. Satosporin C was recovered from the enzymatic reaction by EtOAc extraction and HPLC purification as mentioned previously but with a 20 min 80% MeOH isocratic method. The structure of the compound was confirmed by NMR analysis.

Satosporin C (**3**): colourless oil; $[\alpha]^{25}_{D}$: +23.2 (*c* 0.02, MeOH); IR ν_{max} 2954, 2926, 1735, 1704, 1599. HRESIMS *m/z* 387.2506 $[M+Na]^+$, (calculated for $C_{22}H_{36}O_4Na$ 387.2511, Δ = -0.5 ppm). NMR data are shown in Table 3.4.

3.2.3 Structure elucidation of satosporins

NMR experiments were performed on a Bruker (Milton, ON) 600 MHz spectrometer; samples were dissolved in deuterated methanol (CD_3OD) in a 1.7 mm NMR tube. Spectra were standardized using the proton solvent residual peak signal at 3.31 ppm and carbon signal at 49.00 ppm as a reference[102] and interpreted using the program MestReNova (Mestrelab research, Santiago de Compostela, Spain). The program Xcalibur™ was used to analyse the LC-HRMS data. Infrared spectra were obtained on a Bruker Alpha FT-IR spectrometer. Optical rotation measurements were performed on a Rudolph (Flanders, USA) Autopol III polarimeter in a 50 mm cell containing 1 mL of sample dissolved in methanol at 0.1982%, at 589 nm and 24.7°C. The instrument was verified for its accuracy by using the quartz cell. Average optical rotation was taken from 5 readings.

3.2.4 Determining the relative and absolute stereochemistry by computational modeling

Computational calculations of *in silico* interatomic distances were performed using the PM3 semi-empirical model as implemented in the Spartan 08 package (Wavefunction, Irvine, USA). The nuclear overhauser effect spectroscopy (NOESY) experimental parameters used a mixing time of 500 ms and a relaxation time of 5 ms and the distance of 1.807 angstrom (Å) between the vicinal protons H7a and H7b was used as a reference to calculate the predicted interatomic distances.

In order to predict the theoretical electronic circular dichroism (ECD) spectra of both possible enantiomers of satosporin C, an extensive conformer search over the potential energy surfaces was performed using the PM3 semi-empirical method[103] in conjunction with the automated conformer distribution algorithm implemented in the Spartan 08 package. The theoretical ECD and UV spectra were determined using the time-dependent B3LYP/6-31G*

density functional method[104, 105] incorporating a polarizable continuum solvent model (PCM) representing methanol[106, 107]. Satosporin C (0.5 mg) was sent to Dr Allison Thompson at Dalhousie University and the experimental ECD spectra was obtained on a JASCO (Halifax, NS) J-810 spectropolarimeter in 0.3 mL of HPLC-grade MeOH with a measurement range of 180-700 nm.

3.2.5 Determining the absolute configuration at C16 by the modified Mosher's reaction

Following the method of Latypov et al., the synthesis of the (*S*)-methoxyphenylacetic acid (MPA) ester of satosporin C was achieved by adding a solution of 4-dimethylaminopyridine (DMAP) (10 µg, 0.1 eq) to a 1 mL colourless solution containing satosporin C (0.25 mg, 0.6 µmol), N,N'-diisopropylcarbodiimide (2.0 µl, 25 eq), (*S*)-MPA (0.5 mg, 5 eq) and 3 Å molecular sieves under an N_2 atmosphere[108]. The reaction mixture was concentrated under a stream of N_2 and eluted through a plug of C_{18} using MeOH. The filtrate was evaporated *in vacuo* and further purified by semi-preparative reversed-phase HPLC to tens of micrograms of the desired product which was then analysed by NMR at 25°C and -20°C.

3.2.6 Determining the stereochemistry of the carbohydrate moiety of satosporin A by the modified Tanaka's derivatization

Satosporin A (400 µg) was dissolved in 200 µl of ACN and hydrolysis was performed overnight by adding 200 µl of a trifluoroacetic acid (TFA) solution at a 4 M concentration in order to hydrolyze the sugars. After evaporation, the sample was redissolved in pyridine (1 mL), 2 mg of L-cysteine methyl ester was added and the reaction mixture was heated at 60ºC for 1 h. 2 µL of *o*-tolylisothiocyanate was added and heating was continued for an additional hour and then the product was concentrated *in vacuo*. The residue was suspended in 5% MeOH in H_2O and fractionated over a C_{18} SPE cartridge (500 mg). Five fractions were generated by successive elution with 5% (F1), 25% (F2), 50% (F3), 75% (F4) and 100% MeOH (F5). Samples were dried *in vacuo* and resuspended in MeOH at 0.5 mg/mL and analysed by LC-HRMS using a Thermo Gold C_{18} column and the following solvent system: 20 min isocratic conditions of 87:13 water:ACN followed by a 8 min gradient to 70:30 water:ACN and holding the conditions for 2

min, then a gradient to 100% ACN over 2 min and holding these conditions for 3 min, and finally a return to starting conditions over 2 min for the remainder of the 45-minute run; solvents contained 0.1% formic acid and a flow rate of 300 µL/min was used.

3.2.7 Searching for satosporins in other *Kitasatospora* strains

Eleven cultures from the agriculture research service culture collection (ARS/NRRL) which were labelled as *Kitasatospora* spp. were ordered and cultured as previously mentioned but in 10 mL of production media along with the satosporin producer *K. griseola* MF730-N6 and media blanks; the organisms used in the study are listed in Table 3.1. A simplified extraction procedure was used which consisted of extracting the fermentation with 10 mL EtOAc, allowing 1 h for the fermentations to extract with the solvent at 200 rpm, removal of the organic layer and repeating the process a second time. The EtOAc layers were then combined and dried under air and analysed by LC-HRMS (0.5 mg/mL in MeOH). Retention time and *m*/z of satosporins were used to screen our collection of *Kitasatospora* spp. for the production of compounds 1 and 2.

3.2.8 Biological activity of satosporins

Satosporin A (**1**) was tested following the clinical laboratory standards institute testing standards in a microplate broth assay at concentrations up to 64 µg/mL against the following microorganisms: methicillin-resistant *Staphylococcus aureus* ATCC 33591, vancomycin-resistant *Enterococcus faecalis* EF 379, *Candida albicans* ATCC 14035, *Pseudomonas aeruginosa* ATCC 14210, *Proteus vulgaris* ATCC 12454, *Staphylococcu warneri* ATCC 17917, *Malassezia furfur* ATCC 38593 and *Propionibacterium acnes* ATCC 6919. Satosporin A was also sent to the National Cancer Institute (Bethesda, USA) and tested against 58 cancer cell lines at 10 µM. Finally satosporin A was tested against healthy human keratinocyte (HEKA) and fibroblast (BJ) cell lines at 128 µg/mL and the aglycone (**3**) was tested against methicillin-resistant *S. aureus*, vancomycin-resistant *E. faecalis* and *C. albicans* at 64 µg/mL; these assays were performed by Martin Lanteigne (UPEI) following the clinical laboratory standards institute testing standards (2003) in a microplate broth assay.

Table 3.1. Species of *Kitasatospora* used in the study to assess the potential for satosporin production by members of the same genus.

Species	**Culture collection n°**
Kitasatospora griseola MF730-N6	FERM BP-1045
Kitasatospora griseola	NRRL B-16229
Kitasatospora cystarginea	NRRL B-16505
Kitasatospora grisea	NRRL B-16503
Kitasatospora kifuensis	NRRL B-24284
Kitasatospora mediocidica	NRRL B-16109
Kitasatospora melanogena	NRRL B-16502
Kitasatospora papulosa	NRRL B-16504
Kitasatospora phosalacinea	NRRL B-16230
Kitasatospora streptosporus	NRRL B-16228
Kitasatospora setae	NRRL B-16185
Kitasatospora sp.	NRRL F-6133

3.3 RESULTS AND DISCUSSION

3.3.1 Fermentation of *Kitasatospora griseola* MF730-N6

Following the 24 h incubation of *K. griseola* in the seed medium, growth was evidenced by a cloudy fermentation. After the 48 h shaken incubation in 24 Fernbach flasks containing 500 mL of lean medium, the broths were slightly red in colour and long filamentous mycelia were produced as previously reported in the literature[67, 86]. When all samples were plated to ensure one colony type and thus no contamination of the cultures, small white colonies formed after a 24 h incubation of the lean medium agar plate at 30°C. After 2 to 3 d, the plates still contained a single filamentous colony type where the aerial hyphae had differentiated into spores with a powdery grey surface as expected for this organism[86].

3.3.2 Isolation and structure elucidation of satosporins from *K. griseola* fermentation

From the 12 L of *K. griseola* MF730-N6 fermented in this study, 456.38 mg of crude extract was obtained from eluting the HP20 resin with 100% methanol. The first liquid-liquid partition (H_2O:EtOAC) of this crude extract provided 362.22 mg of product in the EtOAc layer which was then taken through another liquid-liquid partitioning step (hexanes:ACN) to provide 145.77 mg of product in the ACN fraction. Satosporins were obtained from various fractions of the mixture of compounds from the ACN fraction.

Satosporin A

Satosporin A was present in both the ethyl acetate and methanol fractions of the normal-phase flash chromatography performed using a Diol HyperSep™ cartridge. From the 20.17 mg that eluted from the diol column using ethyl acetate, 2.4 mg of satosporin A was obtained (t_R= 20 min) from reverse-phase HPLC using 30:70 water:methanol isocratic conditions for 20 min. Unpure satosporin A was also obtained from the 30% and 50% aqueous ACN fractions of the MeOH fraction mentioned above during reversed-phase flash chromatography using C_{18} HyperSep™. HPLC purification of satosporin A from these fractions was performed as previously mentioned to result in 1.63 mg and 1.29 mg from the 30% and 50% aqueous ACN

fractions, respectively. Fractions from which satosporin A were obtained are illustrated in Figure 3.2.

HRESIMS supported a molecular formula of $C_{34}H_{56}O_{14}$ as the base peak at *m/z* 711.3558 was consistent with the sodium adduct pseudomolecular ion $[M+Na]^+$ (Δ = -0.6 ppm). This also indicated seven degrees of unsaturation in the structure of satosporin A. The ^1H and ^{13}C NMR data, shown in Table 3.2, indicated the presence of carbonyls at δ_C 181.4 (C1) and δ_C 218.1 (C9) respectively in addition to two sugar residues with the characteristic anomeric carbon chemical shifts at δ_C 101.8 (C1') and δ_C 105.1 (C1''). Since the satosporin aglycone was devoid of sp^3 quaternary carbons, the structure of satosporin A was mostly elucidated by the interpretation of the key correlation spectroscopy (COSY) correlations highlighted in Figure 3.3, where cross peaks between hydrogen atoms of adjacent carbons (J^3-coupling) allowed the connectivity of the backbone between carbons 17, 16, 15, 14, 13, 12, 11, 2, 3, 4, 5, 6, 7, 8 and finally back to carbon 3. Furthermore, COSY cross peaks also connected C22 to C16, C21 to C12, C20 to C10 to C11, C18 to C4 and C19 to C6. The remainder of the molecule was elucidated by analyzing the spectrum obtained from a heteronuclear multiple bond correlation (HMBC) experiment which also confirmed the previously assigned connectivities. The partially overlapping signals at δ_H 2.37 (m, H8) and δ_H 2.39 (m, H12) required the HMBC correlations H8/C3, H8/C7, H12/C10 and H12/C14 to unambiguously located the two methines H8 and H12 on the carbon skeleton. Finally, the ketone moiety (δ_C 218.1) was located at C9 due to key HMBC correlations H20/C9, H10/C9, H3/C9, H11/C9, and H8/C9 and the ring closure to form the 8-membered lactone was evident from the key HMBC correlations H2/C1 and H15/C1 (Figure 3.3).

The two sugar residues were identified by interpretation of ^1H-^1H COSY correlations and coupling constant (J) analysis; they revealed the two glycosidic spin systems H1' to H6' and H1'' to H6''. The large coupling constant values for the methine at δ_H 4.39 (d, *J* = 8.0 Hz, H1') and the methine at δ_H 3.47 (t, *J* = 9.0 Hz, H4') placed H1', H2', H3', H4' and H5' in axial positions and suggested that this unit was a *β*-glucopyranosyl residue. This conclusion was in

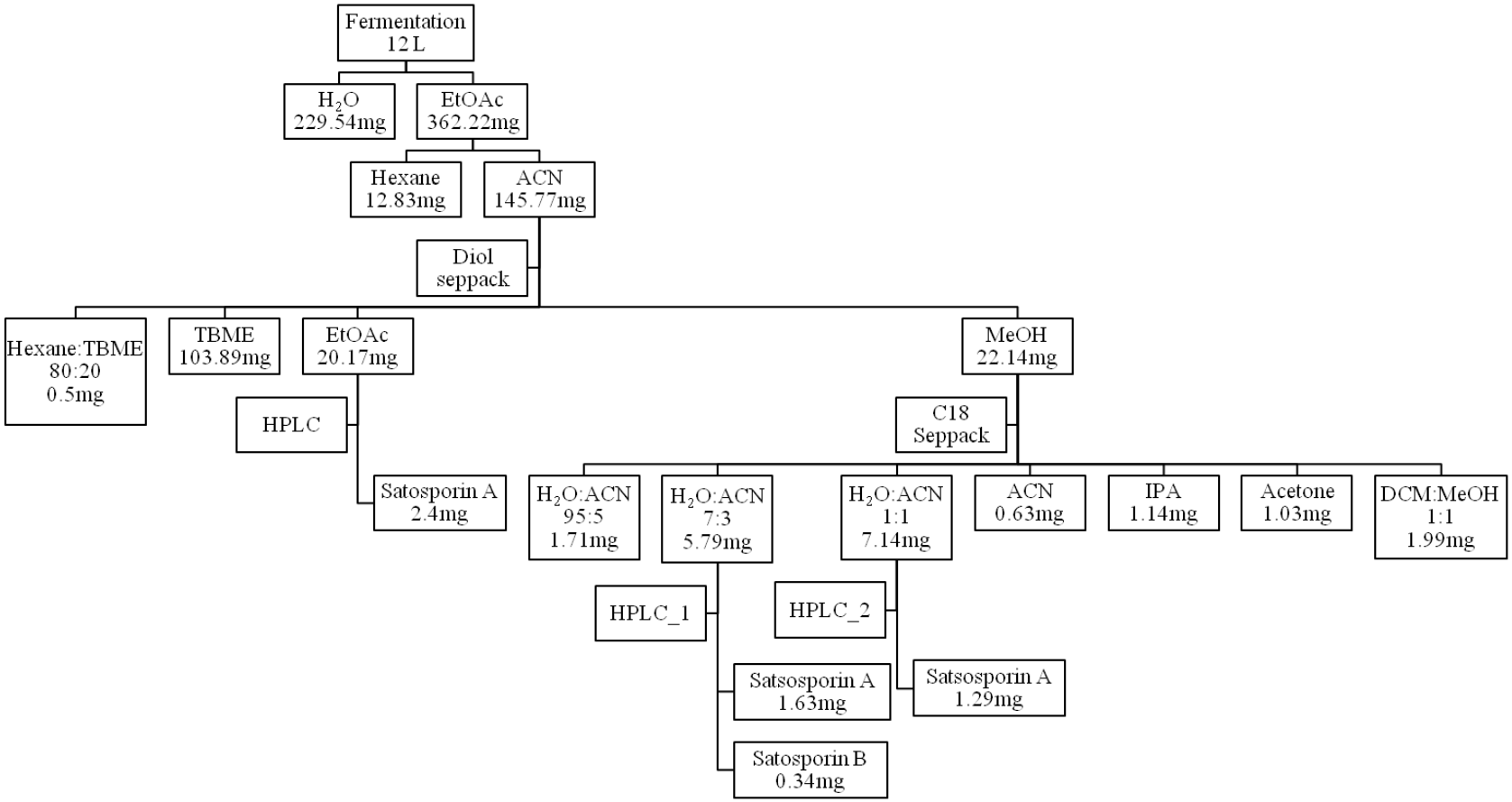


Figure 3.2. Purification scheme leading to the isolation of satosporin A and B from various fractions from the fermentation of 12 liters of *K. griseola*; HPLC, HPLC_1 & HPLC_2: Isocratic water (0.1% formic acid):methanol (0.1% formic acid) 30:70 for 20 min.

Table 3.2. NMR table for satosporin A (**1**); s: singlet, d: doublet, t: triplet, q: quartet, m: multiplet, b: broad, J; coupling constat.

Position	δC, Type	δH (*J* in Hz)	COSY	HMBC	ROESY
1	181.2, C			-	
2	49.7, CH	2.80 t (9.5)	H_3, H_{11}	C_1, C_3/C_4, C_{10},C_{11}	H_4, H_{20}, H_{21}, H_{18}
3	42.0, CH	1.74 ddd (13.1, 9.7, 9.7)	H_2, H_4, H_8	C_1, C_2, C_4, C_7, C_8, C_9, C_{18}	H_{20}, H_{18}, H_{5b}, H_{7b}
4	42.1, CH	1.35 m	H_3, $H_{5a,b}$, H_{18}	-	H_2, H_{11}, H_8
5	44.9, CH_2	1.60 bd (12.6) 0.71 q (12.5)	H_4, H_{5b}, H_6, H_{7a} H_4, H_{5a}, H_6	C_3/C_4 C_3/C_4, C_6, C_7, C_{18}, C_{19}	H_{18} H_3
6	32.0, CH	1.47 m	$H_{5a,b}$, $H_{7a,b}$, H_{19}	-	H_8,
7	36.0, CH_2	2.12 bd (13.2) 0.77 q (12.4)	H_{5a}, H_6, H_{7b}, H_8 H_6, H_{7a}, H_8	C_3 C_6, C_8, C_{19}	H_{19} H_3
8	48.0, CH	2.37 m	H_3, $H_{7a,b}$	C_3, C_7, C_9	H_4, H_6
9	218.1, C			-	
10	43.6, CH	2.55 dq (4.3, 7.7)	H_{11}, H_{20}	C_2, C_9, C_{11}, C_{20}	H_{20}, H_{21}
11	44.6, CH	2.65 m	H_2, H_{10}, H_{12}	C_1, C_2, C_9, C_{10}, C_{12}, C_{13},C_{20}, C_{21}	H_3, H_4, H_{7a}, H_{18}, H_{20}, H_{21}
12	37.6, CH	2.39 m	H_{11}, H_{21}	C_{10}, C_{14}	-
13	40.1, CH_2	1.89 q (15.6, 10.5, 3.1) 0.95 m	H_{14}, H_{12}, H_{13a}	C_{11}, C_{12}, C_{14}, C_{15}, C_{21} C_{12}, C_{14}, C_{15}, C_{21}, C_{22}	H_{15}, H_{20} -
14	39.3, CH	2.34 m	$H_{13a,b}$, H_{15}, H_{22}	C_{15}, C_{16}, C_{22}	-
15	87.1, CH	4.58 d (10.1)	H_{14}, H_{16}	C_1, C_{14}, C_{22}	H_{13a}, H_{21}, H_{22}
16	72.5, CH	4.23 bq (6.3)	H_{15}, H_{17}	C_{17}, $C_{1'}$	H_{14}, H_{22}, $H_{1'}$
17	16.5, CH_3	1.24 d (6.8)	H_{16}	C_{15}, C_{16}	H_{15}, $H_{1'}$
18	19.9, CH_3	0.86 d (6.5)	H_4	C_3, C_4, C_5	H_2, H_3, $H_{4a,b}$, H_{11}
19	22.6, CH_3	0.93 d (6.6)	H_6	C_5, C_6, C_7	$H_{5a,b}$, $H_{7a,b}$
20	15.6, CH_3	1.21 d (6.7)	H_{10}	C_9, C_{10}, C_{11}	H_2,H_3, H_{11}, H_{13a}
21	24.5, CH_3	1.04 d (7.3)	H_{12}	C_{11}, C_{12}, C_{13}	H_2, H_{10},H_{11}
22	18.7, CH_3	0.92 d (6.7)	H_{14}	C_{13}, C_{14}, C_{15}	H_{13a}, H_{15}, H_{16}
*-Glucose (C16)					
1’	101.8, CH	4.39 d (8.0)	$H_{2'}$	C_{16}, $C_{2'}$,$C_{3'}$, $C_{5'}$	H_{16}, H_{17}, $H_{3'}$, $H_{5'}$
2’	74.0, CH	3.44 t (8.5)	$H_{1'}$, $H_{3'}$	$C_{1'}$, $C_{3'}$	-
3’	87.8, CH	3.59 t (8.8)	$H_{2'}$, $H_{4'}$	$C_{2'}$, $C_{4'}$, $C_{1''}$	$H_{1'}$, $H_{5'}$ or $H_{5''}$, $H_{1''}$
4’	69.9, CH	3.47 t (9.0)	$H_{3'}$,$H_{5'}$	$C_{3'}$, $C_{5'}$, $C_{6'}$	$H_{1''}$,
5’	77.2, CH	3.33 m	$H_{4'}$, $H_{6'a,b}$	-	-
6’	62.6, CH_2	3.87 m 3.72 dd (11.7, 5.5)	$H_{5'}$ $H_{5'}$	$C_{4'}$, $C_{5'}$ $C_{5'}$	- $H_{4'}$
*-Glucose					
1’’	105.1, CH	4.58 d (8.1)	$H_{2''}$	$C_{3'}$, $C_{2''}$	$H_{3'}$,$H_{4'}$, $H_{3''}$, $H_{5''}$ or $H_{5'}$
2’’	75.3, CH	3.29 t (9.0)	$H_{1''}$, $H_{3''}$	-	-
3’’	77.6, CH	3.40 t (9.0)	$H_{2''}$, $H_{4''}$	$C_{1''}$, $C_{2''}$, $C_{4''}$	$H_{1''}$
4’’	71.4, CH	3.29 t (8.9)	$H_{3''}$, $H_{5''}$	-	-
5’’	77.7, CH	3.33 m	$H_{4''}$, $H_{6''a,b}$	-	-
6’’	62.5, CH_2	3.89 m 3.64 dd (11.8, 6.3)	$H_{5''}$ $H_{5''}$	$C_{4''}$ $C_{4''}$, $C_{5''}$	- $H_{4''}$

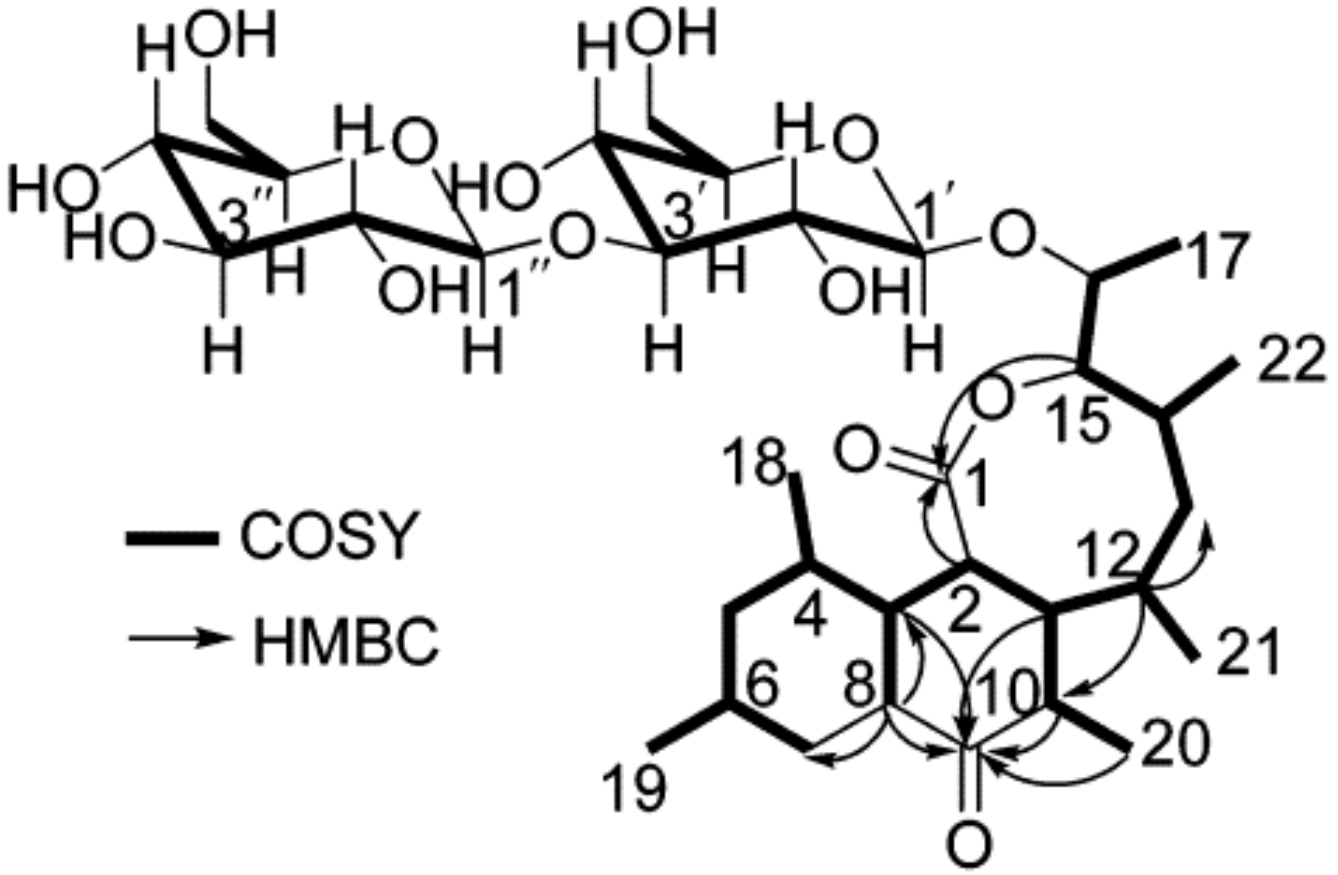


Figure 3.3 Key COSY and HMBC (H→C) of satosporin A.

agreement with the rotating frame overhauser effect spectroscopy (ROESY) data (Figure A.9). The HMBC cross peak between H1' and the resonance at δ_C 72.5 (C16) unambiguously located this sugar on the aglycone. In a similar manner, the second sugar was also identified as a *β*-glucose and was found to be attached to first sugar by a 1,3-*β*-glycosidic linkage according to the key HMBC correlations H1''/C3'and H3'/C1''. The structure of the new polyketide satosporin A (**1**) was thus established (Figure 3.1).

Satosporin B

Satosporin B was purified from a single fraction; the 22.14 mg that eluted with MeOH during normal-phase flash chromatography of the ACN partition was further separated by C_{18} flash chromatography, where satosporin B was purified from the 30% aqueous ACN fraction (5.79 mg) using reverse-phase HPLC and a 70:30 water:MeOH isocratic conditions for 20 min to yield 0.34 mg of satosporin B (t_R=21 min). Due to the strong resemblance of the 1D and 2D NMR data shown in Table 3.3 with those of satosporin A, the structure of satosporin B was rapidly identified as the monosaccharide analogue. Slight changes in chemical shifts were observed in the ^{1}H and ^{13}C spectra at locations 1', 3' and 4' and all signals associated with C1'' to C6'' were absent from spectra of satosporin B, consistent with the only structural difference being the loss of the glucose moiety from C3'. HRESIMS supported a molecular formula of $C_{28}H_{46}O_9$ since the observed base peak at *m/z* 549.3021 was consistent with the sodium adduct pseudomolecular ion $[M+Na]^+$ (Δ = -2.3 ppm). This was also consistent with the loss of a glucose molecule ($C_6H_{10}O_5$) when compared to satosporin A.

Satosporin C

Satosporin C is the aglycone analogue of the new polyketides and although it was never isolated from the producing strain, it was necessary in order to assign the absolute stereochemistry at the C16 stereogenic centre and could not be obtained by acid hydrolysis due to the possible opening of the lactone ring. The aglycone satosporin C was thus obtained by enzymatic deglucosylation of pure satosporin A. Endo-1,3(4)-β-glucanase failed to cleave any

Table 3.3. NMR table for satosporin B (**2**); s: singlet, d: doublet, t: triplet, q: quartet, m: multiplet, b: broad, J; coupling constat.

Position	**δC, Type***	**δH (*J* in Hz)**	**COSY**	**HMBC**	**ROESY**
1	181.4, C				-
2	49.7, CH	2.78 t (9.5)	H_3, H_{11}	C_1, C_{11}	H_4, H_{20}, H_{21}, H_{18}, H_{12Y}
3	42.0, CH	1.74 ddd (13.1, 9.7, 9.7)	H_2, H_4, H_8		-
4	42.0, CH	1.35 m	H_3, $H_{5a,b}$, H_{18}		-
5	44.8, CH_2	1.60 bd (12.8) 0.71 q (12.2)	H_4, H_{5b}, H_6, H_{7a} H_4, H_{5a}, H_6		H_{18}, H_3
6	31.9, CH	1.47 m	$H_{5a,b}$, $H_{7a,b}$,		H_8,
7	35.9, CH_2	2.12 bd (13.9) 0.77 q (12.2)	H_{5a}, H_6, H_{7b}, H_8 H_6, H_{7a}, H_8		H_{19} H_3
8	47.9, CH	2.37 m	H_3, $H_{7a,b}$		H_4, H_6
9	218.6, C				
10	43.6, CH	2.54 dq (4.4, 7.8)	H_{11}, H_{20}	C_9	H_{20}, H_{21}
11	44.5, CH	2.65 m	H_2, H_{10}, H_{12}	C_1,C_{12},	H_3, H_4, H_{7a}, H_{18}, H_{20}, H_{21}
12	37.7, CH	2.39 m	H_{11}, H_{21},H_{13b}		-
13	40.1, CH_2	1.89 ddd (15.6, 10.6, 3.4) 0.95 m	H_{13b},H_{14}, H_{14}, H_{12}, H_{13a}	C_{12}, C_{14},	H_{20} -
14	39.3, CH	2.34 m	$H_{13a,b}$, H_{15},		-
15	87.1, CH	4.58 d (10.4)	H_{14}, H_{16}		H_{13a}, H_{21}, H_{22}
16	72.2, CH	4.23 bq (6.4)	H_{15}, H_{17}	$C_{1'}$	H_{14}, H_{22}, $H_{1'}$
17	16.3, CH_3	1.24 d (5.6)	H_{16}	C_{15}, C_{16}	H_{15}, $H_{1'}$
18	19.9, CH_3	0.86 d (6.6)	H_4	C_3, C_4, C_5	H_2, H_3, $H_{4a,b}$, H_{11}
19	22.5, CH_3	0.93 d (6.4)	H_6	C_5, C_6, C_7	$H_{5a,b}$, $H_{7a,b}$
20	15.6, CH_3	1.21 d (7.3)	H_{10}	C_9, C_{10}, C_{11}	H_2,H_3, H_{11}, H_{13a}
21	24.5, CH_3	1.04 d (7.4)	H_{12}	C_{11}, C_{12}, C_{13}	H_2, H_{10},H_{11}
22	18.7, CH_3	0.92 d (6.8)	H_{14}	C_{13}, C_{14}, C_{15}	H_{13a}, H_{15}, H_{16}
*-Glucose (C16)					
1’	102.1, CH	4.33 d (7.9)	$H_{2'}$	C_{16},	H_{16}, H_{17}, $H_{3'}$, $H_{5'}$
2’	74.6, CH	3.23 t (8.2)	$H_{1'}$, $H_{3'}$		-
3’	78.0, CH	3.38 t (8.8)	$H_{2'}$, $H_{4'}$		$H_{1'}$, $H_{5'}$
4’	71.4, CH	3.32 t (9.2)	$H_{3'}$,$H_{5'}$		
5’	77.7, CH	3.28 m	$H_{4'}$, $H_{6'a,b}$		-
6’	62.6, CH_2	3.87 dd (11.7, 2.0) 3.69 dd (11.7, 5.7)	$H_{5'}$, $H_{6'b}$ $H_{5'}$, $H_{6'a}$	$C_{4'}$, $C_{5'}$ $C_{5'}$	- $H_{4'}$

* Chemical shift were indirectly obtained by interpretation of the heteronuclear single quantum coherence (HSQC) and HMBC data.

glucose molecules from satosporin A, but as seen in Figure 3.4, β-glucosidase from almond was successful , where the rapid disappearance of satosporin A within the first hour of incubation is followed by complete conversion into the aglycone in 8 d. The slow removal of the second glucose moiety may be explained by a greater steric hindrance or a competitive inhibition at the β-glucosidase active site by cleaved glucose molecules[109]. These results also suggested that both sugar moieties might be β-D-glucoses since the enzyme belongs to the enzyme group E.C. 3.2.1.21 which is also known as β-D-glucopyranoside glucohydrolase.

Satosporin A eluted from the column during LC-HRMS analysis and was detected by the mass spectrometer at 3.09 min and the resulting base peak at m/z 711.3573 corresponded to the sodium adduct [M+Na]+ of the disaccharide polyketide (Figure 3.5). Similarly, satosporin B was detected at 3.24 min and the mass spectrum base peak at m/z 549.3033 corresponded to the sodium adduct [M+Na]+ of the monosaccharide polyketide. Finally, the aglycone was detected at 4.12 min and the mass spectrum base peak at m/z 387.2506 corresponded to the sodium adduct [M+Na]+ of the aglycone polyketide. Pure satosporin C was obtained by extracting the incubated reaction with EtOAc and purifying by HPLC; the aglycone eluted at 14.5 min of the 80% MeOH isocratic method.

The structure elucidation of satosporin C was rapidly assessed by comparing the ^{1}H and ^{13}C NMR spectra shown in Table 3.4 to that of satosporin B (**2**). Most signals were similar to those of satosporin B, except that all signals associated with glucose, C1' to C6', were absent from satosporin C spectra and slight changes in chemical shifts at positions 16 and 17 were observed, both of which are consistent with the lack of the carbohydrate moiety at C16. HRESIMS supported a molecular formula of $C_{22}H_{36}O_4$ since the observed base peak at *m/z* 387.2506 was consistent with the sodium adduct pseudomolecular ion $[M+Na]^+$ (Δ = -0.5 ppm). This was also consistent with the loss of a glucose molecule (loss of $C_6H_{10}O_5$) when compared to satosporin B.

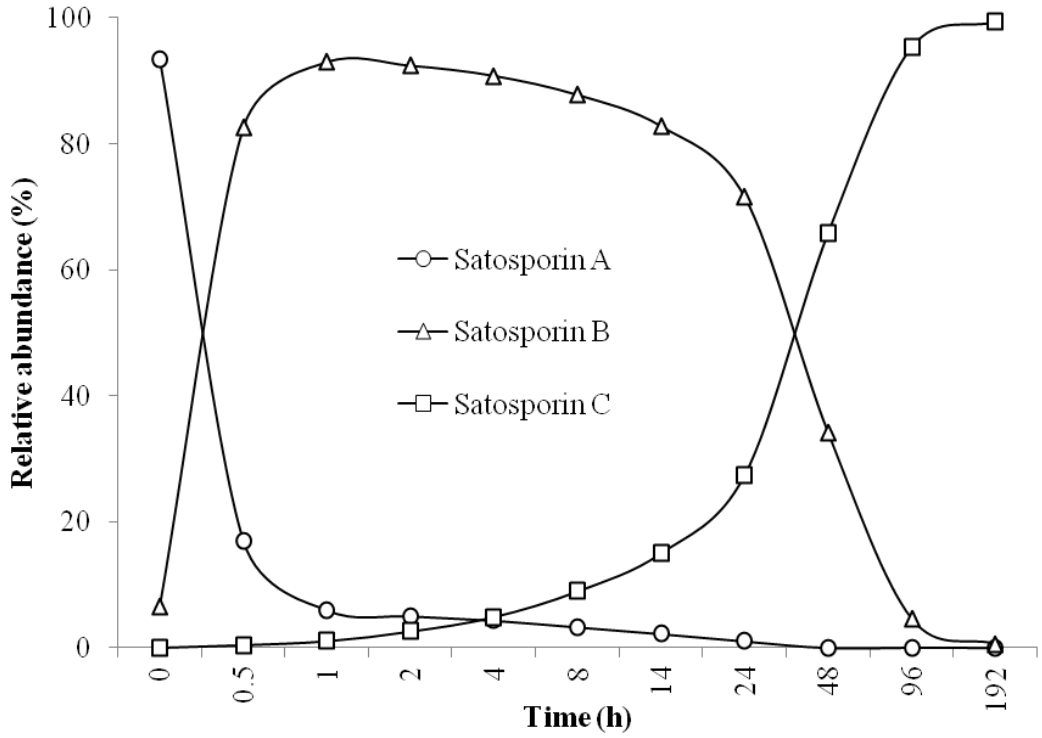


Figure 3.4. Enzymatic deglucosylation of satosporin A into the corresponding monosaccharide and aglycone following an 8 day incubation period with β-glucosidase from almond at 37°C.

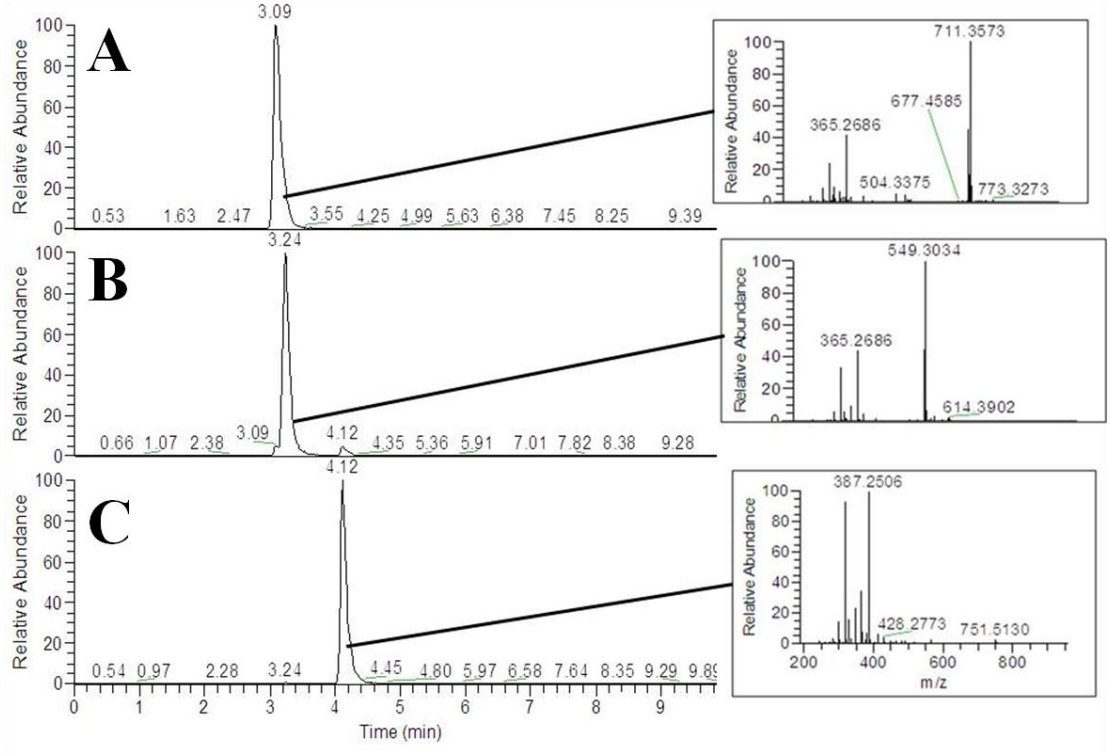


Figure 3.5. Mass chromatogram and corresponding mass spectrum for satosporin A **(A)**, satosporin B **(B)** and satosporin C **(C)**.

Table 3.4. NMR table for satosporin C (**3**); s: singlet, d: doublet, t: triplet, q: quartet, m: multiplet, b: broad, J; coupling constat.

Position	***δ*C, Type**	***δ*H (*J* in Hz)**	**COSY**	**HMBC**
1	181.4, C			-
2	49.6, CH	2.87 t (9.5)	H_3, H_{11}	C1, C_3/C_4, C_{10}
3	42.0, CH	1.71 ddd (13.2, 9.7, 9.7)	H_2, H_4, H_8	
4	42.0, CH	1.36 m	H_{5b}, H_{18}	
5	44.8, CH_2	1.59 bd (13.0) 0.69 q (12.5)	H_4, H_{5b}, H_6, H_{7a} H_4, H_{5a}, H_6	C_3/C_4
6	32.2, CH	1.47 m	H_{5b}, H_{19}	
7	36.1, CH_2	2.12 m 0.75 q (12.2)	H_{5a}, H_{7b}, H_8 H_6, H_{7a}, H_8	
8	48.1, CH	2.36 m	H_3, $H_{7a,b}$	
9	218.8, C			
10	43.7, CH	2.54 dq (4.4, 7.9)	H_{11}, H_{20}	
11	44.4, CH	2.65 m	H_2, H_{10}, H_{12}	C_2, C_9, C_{10}, C_{12}, C_{21}
12	37.6, CH	2.41 m	H_{21}	
13	40.1, CH_2	1.90 ddd (15.6, 10.8,3.8) 0.97 m	H_{13b},H_{14},H_{12}, $H_{14,}$ H_{13a}	C_{11}, C_{14}, C_{15}
14	39.7, CH	2.19 m	$H_{13a,b}$, H_{15}, H_{22}	
15	87.6, CH	4.48 d (10.5)	H_{14}, H_{16}	C_1
16	65.8, CH	4.01 bq (6.4)	H_{15}, H_{17}	
17	20.3, CH_3	1.18 d (6.5)	H_{16}	C_{15}, C_{16}
18	19.9, CH_3	0.85 d (6.5)	H_4	C_3/C_4, C_5
19	22.6, CH_3	0.93 d (6.5)	H_6	C_5, C_6, C_7
20	15.6, CH_3	1.21 d (7.8)	H_{10}	C_9, C_{11}
21	24.5, CH_3	1.04 d (7.5)	H_{12}	C_{11}, C_{12}, C_{13}
22	18.7, CH_3	0.89 d (6.8)	H_{14}	C_{14}, C_{15}

Satosporins possess an unprecedented tricyclic ring system comprising an oxo-decalin unit fused to an 8-membered lactone with 11 stereogenic centres present on the aglycone moiety. This eight membered lactone ring was unusual since most macrolide natural products contain a 14- or 16-membered ring such as in compounds such as erythromycin, bafilomycin and oleandomycin[110-112]. The unique lactone was found in only two other natural products: octalactins, which were isolated from a marine bacterium of the genus *Streptomyces* which showed cytotoxic activity against murine carcinoma (half maximal inhibitory concentration (IC_{50}) =7.2 ng/mL) and human colon cancer (IC_{50}=0.5 µg/mL) cell lines, and gonioheptolides, which were isolated from the bark of the plant *Goniothalamus giganteus* and showed moderate cytotoxicity against various human tumor cell lines with the best median effective dose (ED_{50}) reported as 4.25 µg/mL[113]. The sugar moieties of satosporins were also unusual as most bacterial glycosylated natural products contain 6-deoxyhexoses that may also be further modified[114]. Other glucose containing natural products include vancomycin, chloroeremomycin, rebeccamycin, pyralomycin and glucosylated-oleandromycin [111, 115-117].

3.3.3 Determining the relative and absolute stereochemistry of satosporins

Since the three-dimensional structure of cyclohexane rings has been studied extensively, coupling constants and NOESY correlations were used to determine the relative stereochemistry of the oxo-decalin moiety of satosporins[118]. The apparent doublet of triplets at δ_H 1.71 (H3, J = 13.2, 9.7 Hz) and the two apparent quartets at δ_H 0.69 (H5b, J = 12.5 Hz) and δ_H 0.75 (H7b, J = 12.2 Hz) with large coupling constants placed H3, H4, H5b, H6, H7b and H8 in axial positions on the cyclohexane ring system. These conclusions were further confirmed by the observed nuclear overhauser effect (NOE) correlations between H4/H6/H8 and H3/H5b/H7b (Figure 3.6). The key NOE correlations H8/H11, H2/H4 located H2 and H11on the same face of the carbon skeleton, whereas the methyl group H3-20 was placed on the opposite side, consistent with the NOE correlation H3/H3-20.

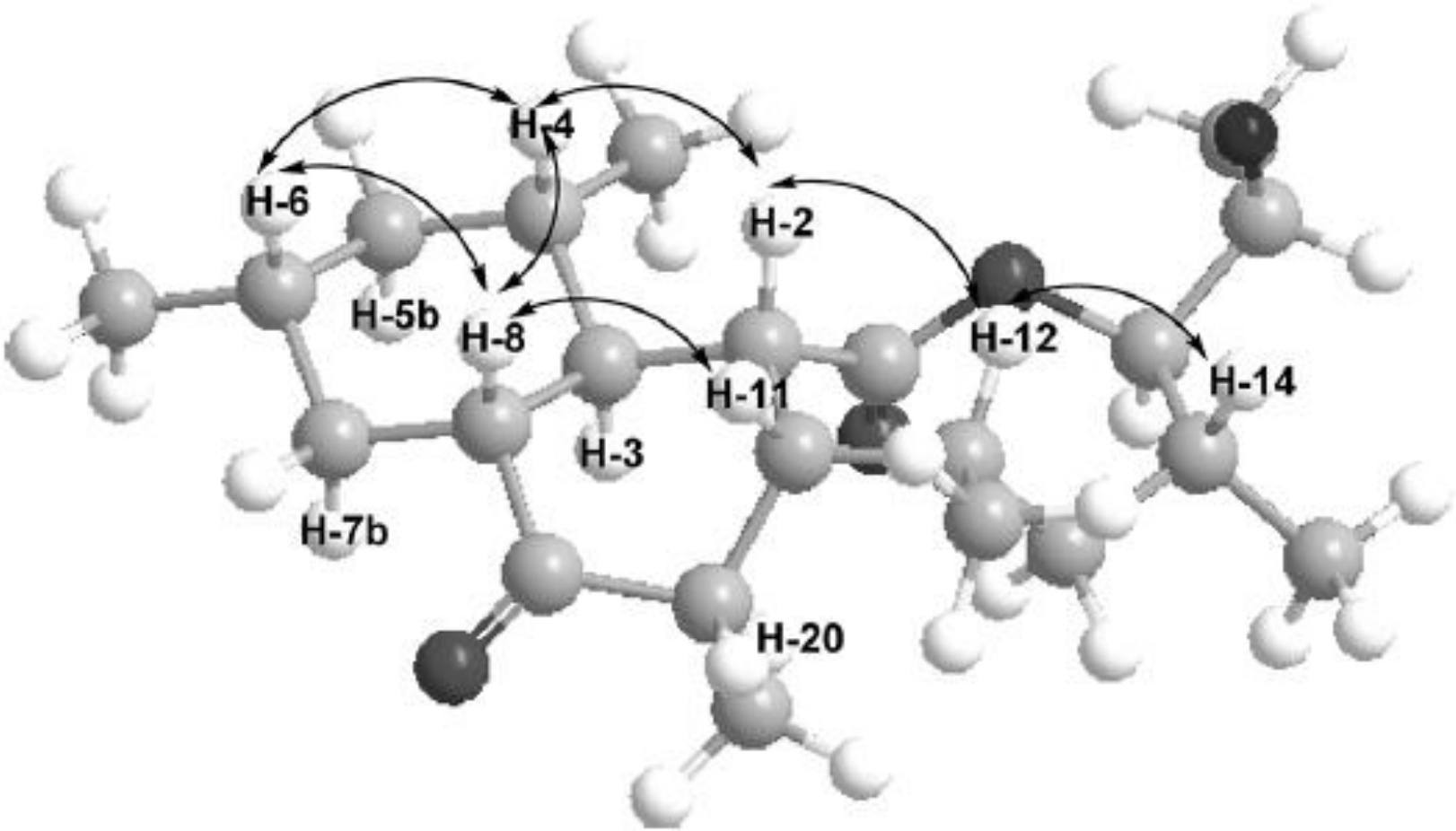


Figure 3.6. Key NOESY correlations of satosporin C.

Since the conformation of the 8-membered lactone was not easily predicted, computational techniques were used to predict *in silico* interatomic distances for each conformer generated by varying the absolute stereochemistry at positions 12, 14, and 15 of satosporin C; this provided the eight conformers listed in Figure 3.7 along with their relative stereochemistry.

In order to simplify the complexity and the amount of computational calculations required to predict the relative and absolute stereochemistries, the aglycone moiety of satosporins was used. This was made possible by the successful purification of the aglycone molecule (**3**) by enzymatic deglucosylation of satosporin A as mentioned previously. Since the presence of the sugars would not affect the three-dimensional structure of the eight-membered lactone, it was possible to extrapolate the results from this experiment to the structures of satosporin A and B. Furthermore, given the stereoselective nature of the domains of modular PKSs, it is not expected that satosporin analogues would have different stereochemistries than satosporin C given that the backbone of the molecule would be made from the same synthases and would only differ in how much post-PKS modifications they underwent by the glycosyltransferase[119]. Analysis of the satosporin biosynthetic gene cluster would confirm whether the origin of satosporin A and B were indeed from the same PKS.

The equilibrium conformer of each isomer was located by scanning the potential energy surface and allowed the prediction of interatomic proton distances. The NOESY experiment allowed the calculation of relative distances between proton atoms by remaining in the linear prediction limit and absolute distances were derived from the known 1.807 Å distance between protons H7a and H7b. This allowed a quantitative comparison between the theoretical structural parameters and the experimental NOE spectrum of satosporin C.

Firstly, three diagnostic distances were used to assess which of the eight stereoisomers matched the experimental results. As illustrated in Figure 3.7 (A), the interatomic distance between H2 and H12 eliminated isomers BAA, BAB, BBA and BBB, while that of H12 to H14

A

		Stereoisomers							
	Experimental (Å)	**AAA**	AAB	ABA	ABB	BAA	BAB	BBA	BBB
Stereochemistry*		*R,R,R*	*R,R,S*	*R,S,R*	*R,S,S*	*S,R,R*	*S,R,S*	*S,S,R*	*S,S,S*
H-2/-H-12	2.456	2.519	2.490	2.498	2.512	~~3.475~~	~~3.599~~	~~3.469~~	~~3.511~~
H-12/H-14	2.187	2.248	2.136	~~3.629~~	~~3.638~~	~~3.657~~	~~3.625~~	2.578	2.248
H-15/H-14	3.030	3.101	~~2.287~~	~~2.337~~	~~2.793~~	3.065	~~2.225~~	~~2.299~~	~~2.987~~

B

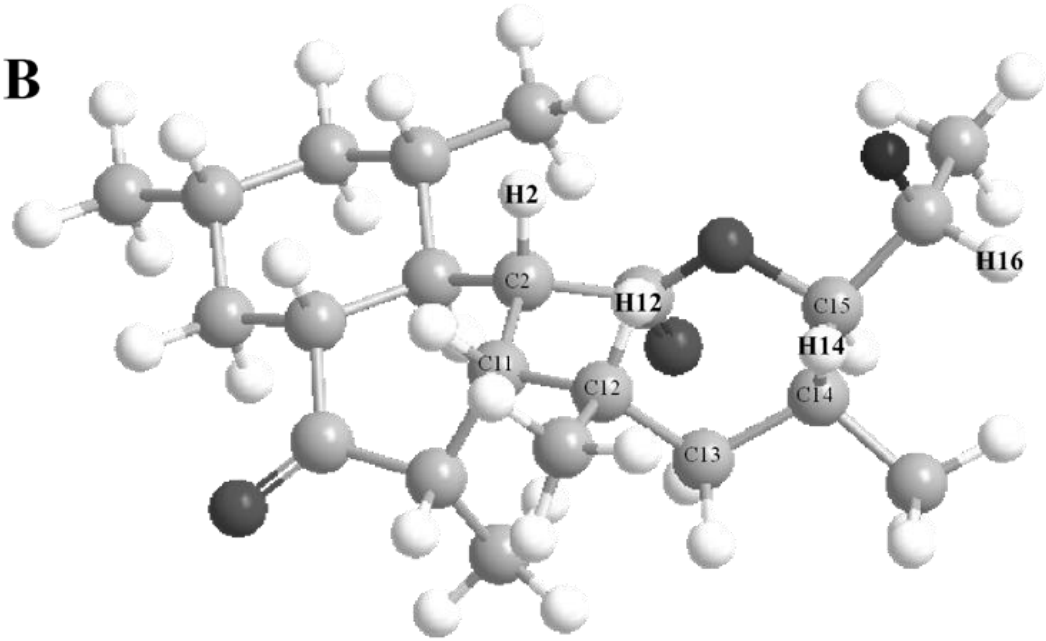


Figure 3.7. Experimental and calculated diagnostic proton distances (Å) for each stereoisomer of satosporin C when varying the relative stereochemistry at carbons 12, 14 and 15 **(A)** and three-dimensional representation of the AAA stereoisomer of satosporin C **(B)**,* refers to the relative stereochemsitry at each stereocentre.

further eliminated isomers ABA and ABB, and finally the H15 to H14 interatomic distance of 3.030 Å further eliminated isomer AAB, leaving AAA (Figure 3.7 (B)), to be the only possible isomer. Secondly, NOE for all interatomic proton distances of AAA were compared to experimental ones; all were consistent with the predicted values of isomer AAA, with the highest amount of divergence being 6.66% as shown in Table 3.5. Furthermore, the strong NOE correlations H2/H12 (2.5 Å) and H12/H14 (2.2 Å) located H12 and H14 on the β side of the molecule, whereas the large coupling constant of the apparent doublet at δ_H 4.48 (H15, J = 10.5 Hz) placed H15 on the α side of the 8-membered lactone ring.

The interatomic distances of H15/H16, H16/H12 and H16/H22 depended on the stereochemistry at C16 and at the time of the experiment this stereochemistry was unknown and thus the distances were not calculated.

Determining the absolute stereochemistry by circular dichroism spectroscopy

The absolute stereochemistry of satosporin C was then determined by comparing the experimental electronic circular dichroism spectrum with theoretically predicted spectra for compound AAA, with *S*, *R* and *R* configuration at carbons 12, 14 and 15 respectively and its enantiomer AAA' with *R*, *S* and *S* configurations at carbons 12, 14 and 15 respectively.

During the extensive conformer search over the potential energy surfaces of AAA and AAA', only two additional structures were located within 3 kcal/mol of the equilibrium conformer and these corresponded to minor rotations of the side chain at position 15. This obviated the usual requirement of assembling the ECD spectra by a Boltzmann average over low energy molecular conformers[120].

The theoretical ECD and UV spectra were determined using the time-dependent B3LYP/6-31G* density functional method[104, 105] incorporating a polarizable continuum solvent model (PCM) representing methanol[106, 107]. Such a methodology has been well-tested and shown to be suitable for such purposes[121-124]. Specifically, the 40 lowest energy transitions (ΔE_i) were calculated along with the corresponding rotatory strengths (R_i) in both the dipole velocity and

Table 3.5. Comparison of proton interatomic distances between experimental NOESY data and predicted values for the stereoisomer AAA of satosporin C. The reference distance of 1.807 angstrom between the vicinal protons H7a and H7b is shown in italics.

			Distance (Å)		Error %
		Integral	Experimental	Theory (**AAA**)	
H15/H22	1	2942206	2.503	2.632	-4.90
H15/H17	2	2729202	2.534	2.656	-4.58
H15/H13a	3	1983040	2.673	2.627	1.75
H15/H14	4	935130	3.030	3.101	-2.30
H15/H16	5	4386154	2.342	N/A	N/A
H16/H15	6	4245438	2.355	N/A	N/A
H16/H17	8	5396434	2.262	N/A	N/A
H16/H22	9	3198782	2.468	N/A	N/A
H2/H18	10	1164252	2.921	2.911	0.35
H2/H4	11	4643264	2.320	2.336	-0.70
H2/H3	12	1323000	2.860	3.056	-6.43
H2/H12	13	3296316	2.456	2.519	-2.50
H2/H11	14	5057380	2.287	2.248	1.73
H11/H2	15	4265480	2.353	2.248	4.66
H11/H21	16	3274562	2.459	2.575	-4.52
H10/H21	17	5563650	2.251	2.140	5.18
H10/H20	18	4218908	2.357	2.368	-0.46
H12/H2	19	3591452	2.421	2.519	-3.89
H12/H14	20	6616316	2.187	2.248	-2.73
H12/H21	21	3917548	2.386	2.457	-2.88
H8/H7a	22	3636900	2.416	2.486	-2.81
H14/H12	23	4991146	2.292	2.248	1.95
H7a/H7b	*24*	*20778476*	*1.807*	*1.807*	*0.00*
H7a/H6	26	2041296	2.660	2.494	6.66
H13a/H13b	28	22097268	1.789	1.797	-0.47
H13a/H20	29	7949812	2.121	2.058	3.05
H3/H20	30	1428382	2.823	2.959	-4.59
H5a/H5b	31	19355970	1.828	1.809	1.08
H5a/H18	32	5145662	2.280	2.391	-4.63
H6/H19	33	3800260	2.398	2.494	-3.83
H4/H18	34	4005418	2.377	2.457	-3.24

dipole length formalisms utilizing the Gaussian 09 (Wallingford, USA) suite of programs. The ECD spectrum was then simulated as a sum of normalized Gaussian functions[120].

We used $\sigma = 0.17$ eV for all transitions but the first two, where we empirically found that $\sigma_1 = 0.05$ eV and that $\sigma_2 = 0.08$ eV produced a significantly better trace. The transition energies, ΔE_i, and rotatory strengths, R_i, are in units of eV and 10^{-40} cgs, respectively. The spectrum may be given in eV but it is more common to present it in nm using the appropriate conversion factor. We also compared the theoretical UV spectra for AAA and AAA' with the experimental UV spectrum in order to determine the so-called UV correction[125]. Due to the fact that a time-dependent density functional theory (TDDFT) generally gives better results for the low-lying states than for the high-energy transitions, where Rydberg states are involved, we determined the UV correction using the spectrum generated from only the first 10 transitions[120]. This corresponded to a UV correction of 12 nm. The UV corrected spectra for AAA and AAA' are shown in Figure 3.8; there was an excellent agreement between the lineshape of the experimental spectrum and that of AAA. In particular, the direction of all Cotton effects are reproduced exactly, though the peaks near 250 nm and 300 nm are not as broad as those in the experimental spectrum. The ECD spectrum of AAA' is, as expected, opposite in sign to that of AAA and thus a clear absolute stereochemical assignment was made. The observed negative Cotton effect at 295 nm in the spectrum of satosporin C, corresponding to the $n \rightarrow \pi^*$ transition of a carbonyl group, along with the remaining data collectively indicated that the configuration of satosporin C was 2*R*,3*R*,4*S*,6*R*,8*R*,10*S*,11*R*,12*S*,14*R*,15*R*.

3.3.4 Determining the absolute configuration at C16 by the modified Mosher's reaction

The configuration of the C16 stereogenic centre was determined from interpretation of NOESY data and further confirmed by chemical derivatization of the secondary alcohol in satosporin C (3) using a modified Mosher methodology. Indeed, the strong NOE correlations H15/H16 (2.342 Å) and H15/H17 (2.534 Å) indicated that H15 and the hydroxyl group held an anti conformation while the NOE correlation H16/H3-22 (2.468 Å) suggested an *R* configuration

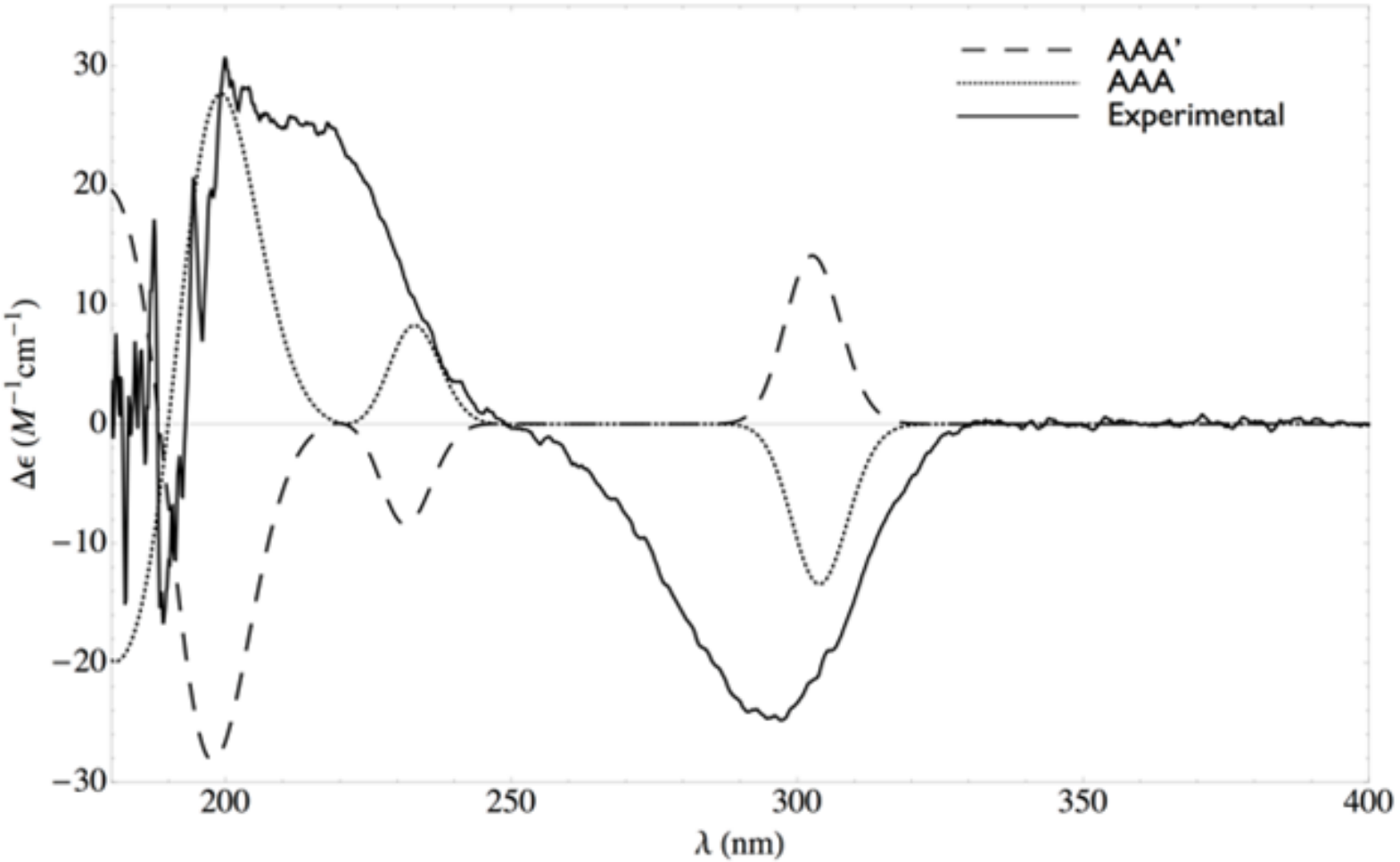


Figure 3.8. Calculated (dashed lines) electronic circular dichroism (ECD) spectra for both possible enantiomers AAA and AAA' and experimental (solid line) ECD spectrum obtained from satosporin C sample.

for C16. This observation was further validated by the synthesis of the (*S*)-methoxyphenylacetic ester of satosporin C and analysis of ^{1}H NMR spectra (Figure 3.9) recorded at two different temperatures: T_1: 25 °C and T_2:-20 °C. The difference of chemical shift $\Delta\delta^{T_1,T_2}$ being negative for ^{1}H signals H2, H8, H11, and H15 indicated that the cyclic portion of the aglycon was shielded by the phenyl ring in the *ap* conformer, and consequently confirmed the *R* configuration at C16.

3.3.5 Determining the stereochemistry of the carbohydrate moiety of satosporin A by the modified Tanaka's derivatization

In order to determine which glucose enantiomers were present in satosporin A, a modified Tanaka derivatization was performed on the hydrolyzed sugars using pyridine, L-cysteine methyl ester and *o*-tolylisothiocyanate to give rise to diastereomeric sugar derivatives which would be distinguishable via LCMS analysis and compared to derivatized sugar standards.

Screening for *m*/*z* of 447.1252±5 ppm, which corresponded to the $[M+H]^+$ adduct of the isothiocyanate derivatives (Figure 3.10), revealed that the standard and polyketide glycoside derivatives were in fraction F3 during reverse-phase chromatographic separation. The L-glucose derivative standard was detected at 14.53 min, while its enantiomer's, D-glucose, derivative was detected at 16.40 min. The derivatized sugars hydrolyzed from satosporin A were detected by the mass spectrometer at 16.48 min (Figure 3.11); since the retention time was the same as that of the derivatized D-glucose standard, sugars of satosporin A were determined to both be D-glucose.

Although the experiment was not performed on the sugar moiety of satosporin B, it is unlikely that it would contain L-glucose given that the glycosyltransferase enzyme presumably used the monosaccharide as a substrate to attach the second sugar. The examination of the biosynthetic gene cluster for satosporins in *K. griseola* would provide further evidence as to whether a single enzyme or two different enzymes are involved in the glycosylation steps leading to the biosynthesis of satosporins.

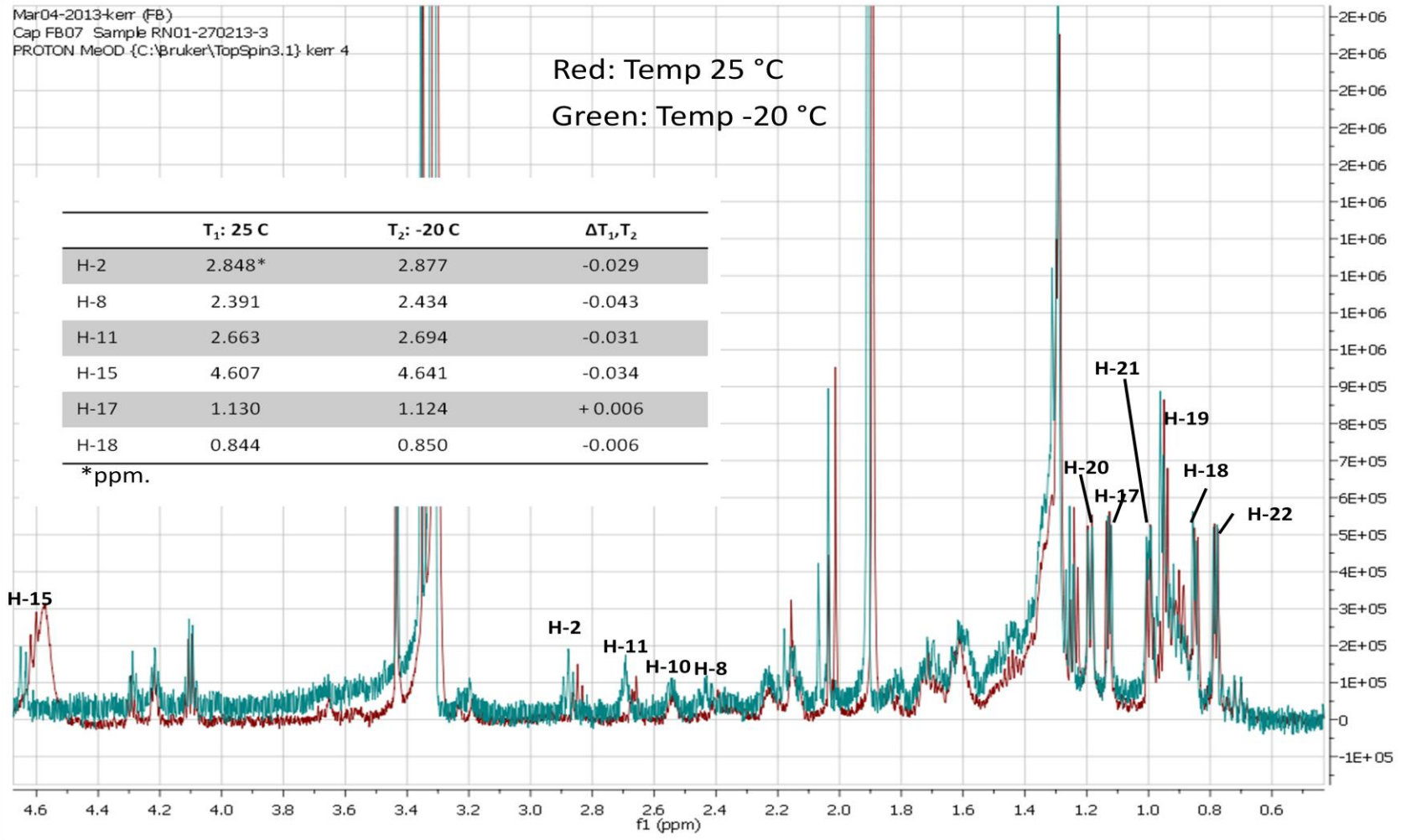


Figure 3.9. Overlay of ^{1}H NMR spectra of the (*S*)-methoxyphenylacetic ester of satosporin C, obtained at 25°C and -20°C, to determine the absolute stereochemistry at C16.

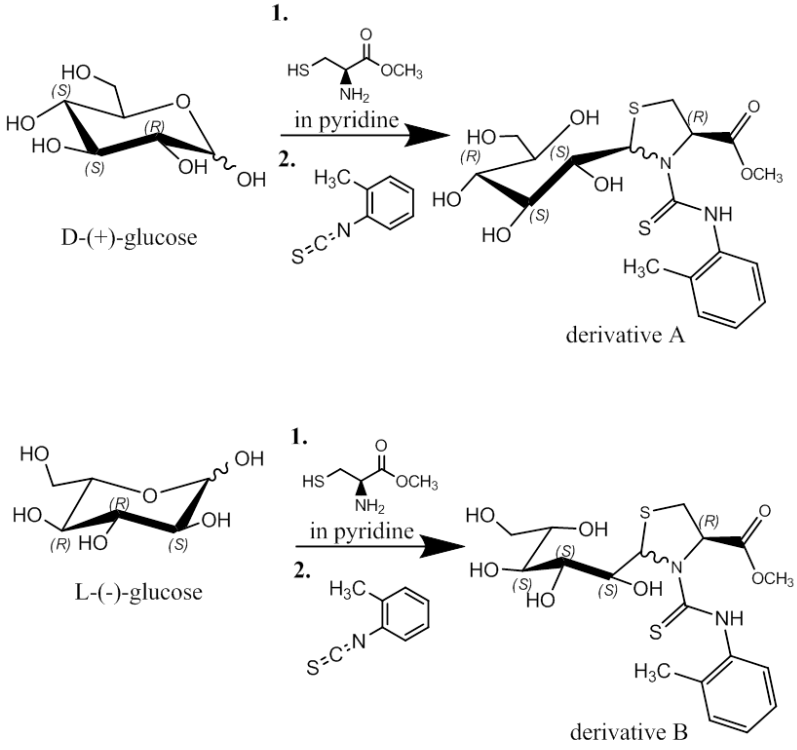


Figure 3.10. Sugar derivatization reaction using first L-cysteine methyl ester in pyridine and *o*-tolylisothiocyanate to give rise to the diastereomeric derivatives.

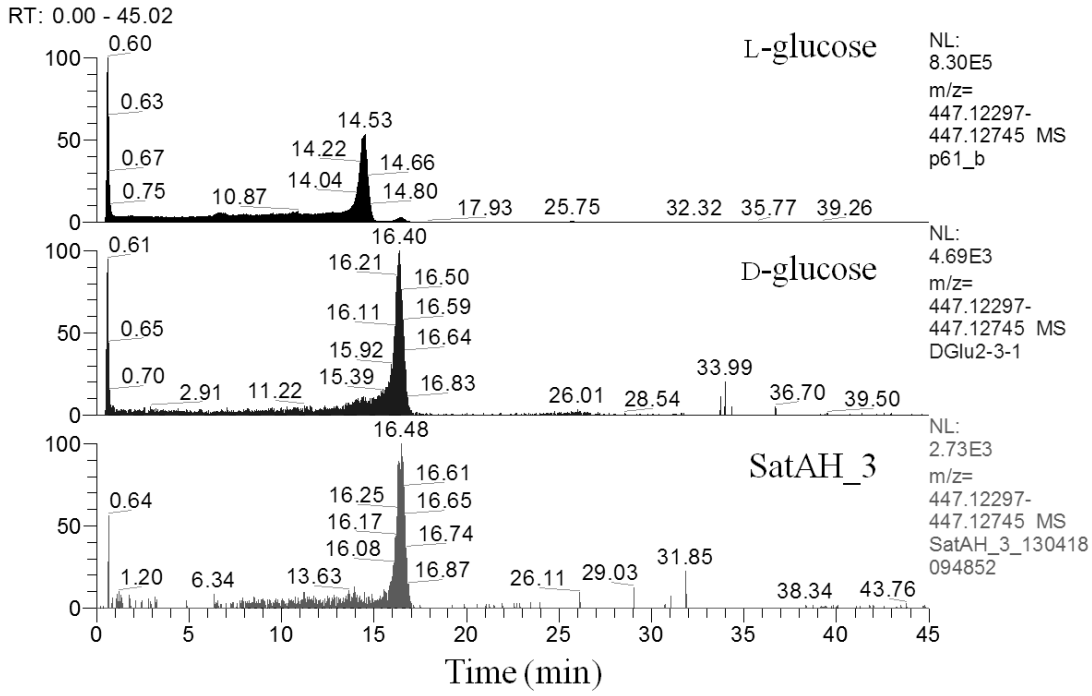


Figure 3.11. HRMS chromatograms for *m*/*z* of 447.1252±5 ppm showing the retention times associated with both glucose enantiomer derivative standards (black) compared with the retention times of the sugars from the polyketide disaccharide, satosporin A (grey).

3.3.6 Searching for satosporins in other *Kitasatospora* strains

Apart from the current producer *K. griseola* MF730-N6, no other *Kitasatospora* strain investigated here was capable of producing satosporins, including a different strain of the same species, *K. griseola* NRRL B-16229. Until the same or a very similar gene cluster is found in another organism, it is difficult to comment on the evolutionary history of this gene cluster, but it is clear that it was acquired from *K. griseola* MF730-N6 late in the diversification of this genus given that no other *Kitasatospora* strain produced the compound. This is in contrast to bafilomycins, which can be isolated from over 2% of *Streptomyces* and some *Kitasatospora* and is thus thought to have been acquired earlier in the evolutionary history of actinomycetes[28]. Terpentecin is another compound which seems to be produced by *K. griseola* MF730-N6 but not by *K. griseola* NRRL B-16229. Terpentecin has, however, been isolated from *Streptomyces* sp. s464[126], which may be another example of convergent evolution, where the production of the same compound has been adopted by different microbial lineage to provide an evolutionary advantage to the producing organism. It is also possible that *Streptomyces* sp. s464 may be more closely related to *K. griseola* as anticipated since members of the two genera are difficult to differentiate and could thus have been mislabelled as *Streptomyces*[97]. A more comprehensive genetic analysis of the secondary metabolite gene clusters present in the genomes of *Kitasatospora* and *Streptomyces* species would be necessary in order to gain knowledge on the evolutionary origins of the satosporin biosynthetic gene cluster.

3.3.7 Biological activity of satosporins

Satosporin A showed no inhibitory activity against methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant *Enterococcus faecalis*, *Candida albicans*, *Pseudomonas aeruginosa*, *Proteus vulgaris*, *Staphylococcus warneri*, *Malassezia furfur* and *Propionibacterium acnes* at 64 µg/mL or against the National Cancer Institute's 58 cancer cell lines at 10 µM. It was however also non-toxic to the healthy human keratinocyte (HEKA) and fibroblast (BJ) cell lines at 128 µg/mL, providing a promising perspective for future applications of this natural product. Since the production of complex compounds such as satosporins is

energetically costly to the producing bacterium, it is strongly believed that all natural products serve a biological function that benefits the host[1]. The pharmaceutical and cosmetic industries have taken advantage of these biological activities and can develop the natural product into drugs such as antibiotics, antifungal agents, anticancer drugs etc. Since no activity has been described for satosporins, we are unable to make predictions as to the possible human and animal health applications these compounds may have. We can however conclude that whatever biological activities satosporins may have, there is a good chance that the activity will be specific towards the target and, most importantly, it appears to be non-toxic to healthy human cells.

3.4 CONCLUSION

This research described the isolation and structure elucidation of two new polyketides, satosporin A and B, isolated from fermentations of *K. griseola.* Along with the uncommon glucosylation of the polyketide aglycone, the eight-membered lactone ring seen in satosporins is rare amongst natural products; the compounds octalactin and gonioheptolides are the only other natural products with this saturated octalactone. Fermentations of *K. griseola* MF730-N6 is thus an efficient means to obtain these rare chemical entities.

As of 2012, twelve species were described as belonging to the genus *Kitasatospora* and almost half of them were also reported to produce a natural product[86]. Efforts to selectively isolate members of this genus have been fruitful in describing new species as well as new compounds with interesting chemistry and bioactivity[98, 127]. This research supports previous findings that members of the genus *Kitasatospora* are prolific producers of secondary metabolites and that further investigations of their metabolome may provide additional novel natural products. The biosynthetic scheme leading to the production of satosporins will be discussed in the following chapter.

CHAPTER 4 - ORGANIZATION OF THE SATOSPORIN BIOSYNTHETIC GENE CLUSTER IN *KITASATOSPORA GRISEOLA*

4.1 INTRODUCTION

Polyketides are an important class of natural products with major roles in human therapeutics as antimicrobials, antifungals, immunosuppressants, cholesterol-lowering and anticancer agents[128]. Although the majority of polyketides are biosynthesized from very few starter and extender building blocks (i.e. acetate, propionate, malonyl-CoA and methymalonyl-CoA), the tremendous structural diversity within this class is a result of the vast combinations of reducing domains within each module of the PKS. For every extender unit, the β-ketone may be sequentially reduced to an alcohol, an alkene or an alkane via a ketoreductase, dehydratase and enoylreductase domains, respectively[9]. Since analysis of the protein sequence of PKSs reveals which domains are present within each module, the ketide product of Type I PKSs can easily be predicted from genetic information.

Satosporins A (**1**) and B (**2**) (Figure 4.1) are two polyketides recently isolated from the terpentecin-producing actinomycete *Kitasatospora griseola* (see Chapter 3) with unique structures featuring a tricyclic ring system comprising an oxo-decalin ring fused to an octalactone to which either one or two glucose units are attached[129]. Their biosynthesis (Figure 4.2) was predicted to occur via multimodular type I PKSs containing one loading and seven extender modules incorporating a rare lactate starter unit along with five methylmalonate and two malonate extender units, resulting in a 17-carbon ACP-bound ketide. Following macrocyclization and release from the synthase via the thioesterase domain, the macrolide was predicted to then undergo intramolecular cyclization via the removal of the α-proton followed by the nucleophilic attack to the β-carbon forming a new σ-bond from π electrons resulting in a hydroxyl leaving group and yielding the aglycone. The addition of a glucose molecule via a glycosyltransferase would result in satosporin B (**2**), while the transfer of two glucose molecules would yield satosporin A (**1**).

Although the biological activity of these compounds remains undetermined, insight into the genes that govern its biosynthesis can lead to the generation of bioactive analogues while

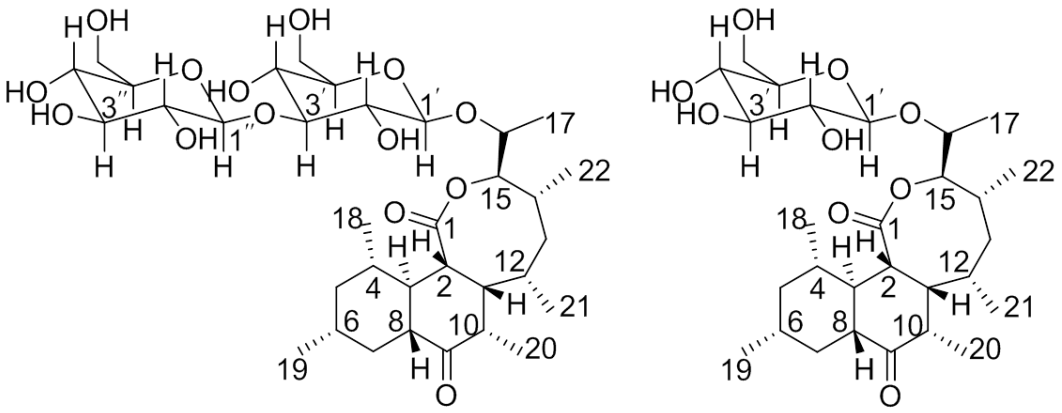


Figure 4.1. Structure of the polyketides satosporins A and B from *K. griseola* MF730-N6.

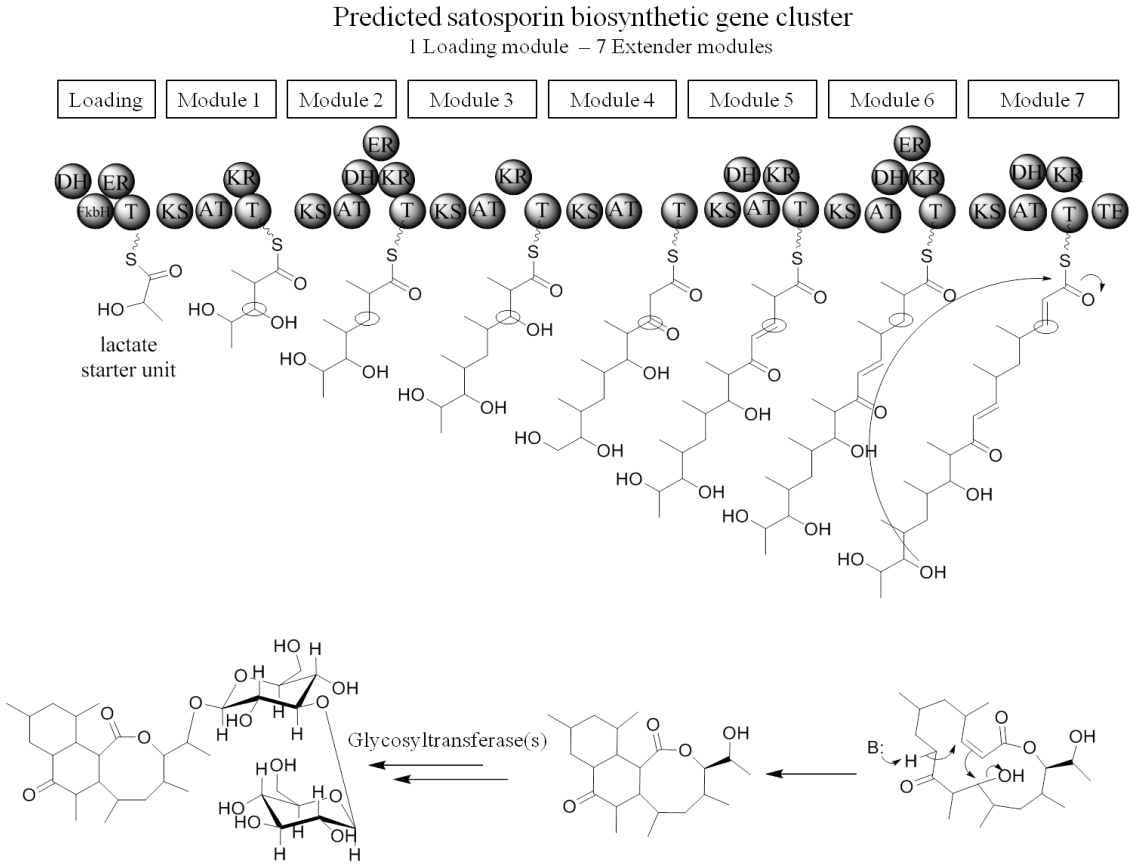


Figure 4.2. The initially proposed structure-based satosporin biosynthesis involved a rare, but documented, lactate starter unit, a ketone originating from the lack of β-keto processing enzymes at module 4, an intramolecular cyclization of the macrolactone to yield the aglycone and finally one or two glycosyltransferases to glucosylate the aglycone; circles indicate the resulting functional group from the domains present within the module. AT: acyltransferase, KS: ketosynthase, ACP: acyl carrier protein, KR: ketoreductase, DH: dehydratase, ER: enoylreductase, TE: thioesterase.

simultaneously increasing the knowledge about polyketide biosynthetic genes to improve functional annotation predictions. The objective of this study was thus to locate and annotate the gene cluster from *Kitasatospora griseola* MF730-N6 responsible for the biosynthesis of satosporins.

4.2 MATERIALS AND METHODS

4.2.1 Fermentation of *Kitasatospora griseola* and chemical extraction

Two seed cultures of *K. griseola* MF730-N6 were grown, first from a scraping from the sporulating organism on an agar plate and then using a 5 % inoculum of the first seed, in disposable culture tubes containing 10 mL seed medium (10 g/L yeast extract and 10 g/L glucose) for 48 h each at 30°C and 200 rpm. 10 mL of lean medium[67] containing 0.4 g glucose, 0.8 g galactose, 0.8 g maltose, 1.6 g dextrin, 0.8 g Bacto™-Soytone and 0.3 g $(NH_4)_2SO_4$ per liter was then inoculated using 500 µL from seed #2 and allowed to grow for 48 h under the same conditions. Purity plates were prepared following fermentations to ensure one colony type and thus no contamination. To assess satosporin production, the fermentations were extracted by shaking twice with 10 mL ethyl acetate (EtOAc) for 1 h and drying the combined organic layers under air. The crude extract was then analysed by liquid chromatography-high resolution mass spectrometry (LC-HRMS).

4.2.2 Sequencing of the *sat* cluster

Polyketide KS-AT region amplification using degenerate primers

Amplicons were generated with EconoTaq® 2X plus green (Lucigen) master mix using the previously developed primer pair K1F and M6R and using *K. griseola* genomic DNA as a template with the following cycling conditions: an initial denaturation at 94°C for 2 min followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 59.4, 57.7, 54.6, 52.4, 50.9 and 50°C for 15 s and polymerization at 72°C for 1:20 min and a final extension at 72°C for 5 min. Amplicons were then gel-purified and cloned into *Escherichia coli* using the instaTAclone

PCR cloning kit (Fermentas, Thermo Scientific™) following the manufacturer's instructions. Ampicillin resistant clones were then subcultured and either subjected directly to colony PCR, as described in Chapter 2, or their pDNA extracted and used as template for PCR using M13F and M13R primers[130] using the cycling conditions mentioned above except that the annealing temperature was 45°C. Amplicons were then subjected to restriction analysis by digesting 20 µL of the PCR reaction with XmaI and RsaI enzymes according to the manufacturer's instructions. Following electrophoresis, banding patterns were compared and grouped by similarity. All amplicons were then sent for sequencing with M13F primer and their DNA sequence was used to dereplicate identical amplicons using the program *Contig*Express® (Vector NTI Advance™, Life Technologies) and also to query a local bafilomycin sequence databases using the nucleotide Basic Local Alignment Search Tool (BLAST®n,) using the freeware BioEdit (Carlsbad, USA) sequence alignment editor program[131]. Primers and strains used throughout this study are listed in Table 4.1.

Illumina® sequencing

Genomic DNA was isolated from cultures of *K. griseola* MF730-N6 grown for 48 h in tryptic soy broth (17 g pancreatic digest of casein, 3 g Bacto™-Soytone, 5 g NaCl, 2.5 g K_2HPO_4 and 2.5 g dextrose per liter and pH adjusted to 7.3) at 30°C and 200 rpm using the PureLink genomic DNA mini kit (Invitrogen). The sample was sent to Dr Nathan Magarvey at McMaster University for Illumina® sequencing using MiSeq sequencer and data was provided as contigs assembled from 150 bp reads and visualized in Geneious (Biomatters, San Francisco, USA). Nodes greater than 5 kb and contigs larger than 1 kb were submitted to antiSMASH at the following URL: http://antismash.secondarymetabolites.org/[38]. Sequences containing polyketide biosynthetic genes were then submitted to BLAST®p to query the National Center for Biotechnology Information (NCBI) database following ORF prediction using FramePlot 4.0 at the following URL: http://nocardia.nih.go.jp/fp4/[132].

Table 4.1. Primers and plasmids used during this study.

Primer name	Sequence or relevant characteristics[a]	Reference
M13F	TGTAAAACGACGGCCAGT	[130]
M13R	CAGGAAACAGCTATGACC	[130]
K1F	TSAAGTCSAACATCGGBCA	[133]
M6R	CGCAGGTTSCSGTACCAGTA	[133]
N2491KO_L_F	GCTTCGGAATTCCGAGCAGTCCGAGGTCTTCC	This study
N2491KO_L_R	GTAGCATGCGTCCACGCTGTCCTCGTCGG	This study
N2491KO_R_F	GTAGCATGCGGCGAGGACCAGGTGGCGAT	This study
N2491KO_R_R	GACTGCAAGCTTGAAGAACGCCGCGTCGAAGT	This study
N2491_confirm_F	CGTGGTGAACGGACCCGCGG	This study
N2491_confirm_R	CCTTGGCCGTGACGACGTCG	This study
pSET151	Th^R, Apr^R, *xyl*E, rep^{pUC}, *ori*T, *lac*Zα	[69]

[a] Th^R:thiostrepton resistance marker, Apr^R: apramycin resistance marker, *xyl*E: catechol 2,3-dioxygenase reporter gene, rep^{pUC}: *E. coli* replication function from pUC, *ori*T: origin of transfer from RK2 and *lacZα*: β-galactosidase for blue-white screening.

PacBio sequencing

Genomic DNA was isolated from 10 mL *K. griseola* cultures grown overnight in seed medium containing 10 g/L glucose and 10 g/L yeast extract at 30°C and 200 rpm using the Qiagen genomic tip 100/G kit. The sample was kept at 4°C and 500 µL of 80 ng/µL gDNA was sent the following day to the McGill University and Génome Québec Innovation Centre to be sequenced using the Pacific Biosciences® RS II platform; data was provided as contigs from a *de novo* assembly and visualized in Geneious.

4.2.3 Insertional inactivation of module 1

Module 1 of the putative satosporin biosynthetic gene cluster in quiver_44 was inactivated through the insertion of the apramycin resistance gene via double crossover of homologous regions[71]. Briefly, two ~1.5 kb regions of homology were PCR amplified from *K. griseola* gDNA using primers pairs N2491KO_left_F and N2491KO_left_R, N2491KO_right_F and N2491KO_right_R using EconoTaq® 2X plus green and the following cycling conditions: an initial denaturation at 94°C for 2 min followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 72.3°C for 15 s and polymerization at 72°C for 1:30 min and a final extension at 72°C for 5 min. This gave rise to amplicons N2491KO_left (bp: 12-1491) and N2491KO_right (bp: 1519-2889), which following column purification were restricted with EcoRI-HF/SphI-HI and SphI-HF/HindIII-HF, respectively, to generate the proper sticky ends required for the assembly of the deletion plasmid. The apramycin resistance gene was used as a selection marker and was obtained as described previously in Section 2.2.4. Plasmid DNA from pSET151[69] was restricted with EcoRIHF/HindIII-HF and the 6 kb band was gel-purified.

All components of the deletion plasmids (left arm, right arm, apramycin resistance gene and the plasmid pSET151) were added to a microcentrifuge tube along with T4 DNA ligase and the reaction components according to the manufacturer's instructions. 5 µL of the ligation reactions were then used to transform chemically competent *E. coli* and plated on LB containing 50 µg/mL apramycin. Colony PCR was first performed on *E. coli* harbouring pN2491KO using

primers N2491KO_left_F and N2491KO_right_R following the same cycling conditions as before except that the polymerization time was extended to 3 min to generate 3819 bp screen_N2491KO amplicons. Plasmid DNA was then extracted from transformed colonies and restriction digests were performed using SacI/PstI and AatII.

The remainder of the knock-out procedure was as described in Section 2.2.4 and PCR was used to confirm the genotype of *K. griseola* ΔN2491 using genomic DNA as template, the new primer pair N2491_confirm_F/R (which annealed within the homologous arms) and similar cycling conditions as before except that the annealing temperature was 68.6°C and the polymerization time was 2:30 min. Genomic DNA was obtained by phenol chloroform extraction[71]. Production of satosporin in wild-type and knock-out strains was assessed by culturing and extracting the fermentations as described in 4.2.1 and assessing whether or not satosporins were present in the crude extracts by LC-HRMS.

4.2.4 Sequence analysis of the sat cluster

The DNA sequence of quiver_44 was submitted to the antiSMASH tool online. Open reading frames (ORF) and ribosome binding sites (RBS) were confirmed using FramePlot 4.0 and annotation was confirmed by BLAST® analysis against publicly available databases. The NRPS-PKS and Dock_Dom_Anal interfaces of the structure based sequence analysis of polyketide synhtases (SBSPKS) program were also used to analyse the PKSs[134]. These sequences have been submitted to GenBank under accession number JXZB00000000.

4.3 RESULTS AND DISCUSSION

4.3.1 Fermentation of *Kitasatospora griseola* and chemical extraction

Following a 48-hour fermentation, growth of the organism in the lean medium was evidenced by a cloudy reddish medium and formation of aerial mycelia as expected for this organism[86]. Purity plates contained a single phenotype where white colonies appeared after 24 h incubation and no other colony type appeared over the next 3 d. Following extraction by EtOAc and LCMS analysis, satosporin was detected in crude extracts, as reported previously[129],

evidenced by the presence of the pseudomolecular ion $[M+H]^+$ of satosporin A (**1**) and its sodium adduct with *m*/*z* 689.3749 and 711.3564 (Δ = 0.9 ppm and 0.2 ppm), respectively, in the mass chromatogram (Figure 4.3).

4.3.2 Sequencing of the sat cluster

Degenerate primers to amplify KS clones

The first strategy to elucidate the satosporin biosynthetic gene cluster was to use previously designed K1F and M6R degenerate primers designed to anneal to conserved regions of the ketosynthase and methylmalonyl-CoA transferase modules, respectively[133]. The PCR reaction using these primers and *K. griseola* genomic DNA generated amplicons roughly 1200 bp long. Following gel-purification and cloning of the amplicons into *E. coli*, inserts were amplified by PCR using M13 forward and reverse primers. The amplicons were then subjected to a restriction digest with XmaI and RsaI to give rise to the fragmentation patterns shown in Figure 4.4. Fifty-eight amplicons were categorized into 11 groups based on similarity of the restriction pattern as listed in Table 4.2. From this analysis, two thirds of the amplicons analysed were grouped into categories A, B and C; amplicons in category E were not necessarily identical as the amplicons were not restricted by the enzyme and, thus, provided no information regarding the ~1200 bp ketosynthase to methylmanolyl-CoA transferase (KS-AT) region.

Since the restriction analysis provided no information on the actual DNA sequence of the amplicon, each was also sent for sequencing. First, an assembly was performed to dereplicate identical sequences. No sequence data was obtained for four of the amplicons using M13F primer, 47 sequences fell into 6 contigs and 7 sequences did not form any contig. The analysis provided similar groupings to those obtained by restriction pattern analysis, where the majority of sequences fell into two groups, 1 and 4 of Table 4.3. The smaller groups were less consistent between the two methods, which was expected given the high resolution of the sequence assembly and the much lower resolution of the restriction analysis; the latter was however inexpensive and therefore would be an appropriate dereplication tool for large datasets where only a few representatives from each group would need to be sequenced.

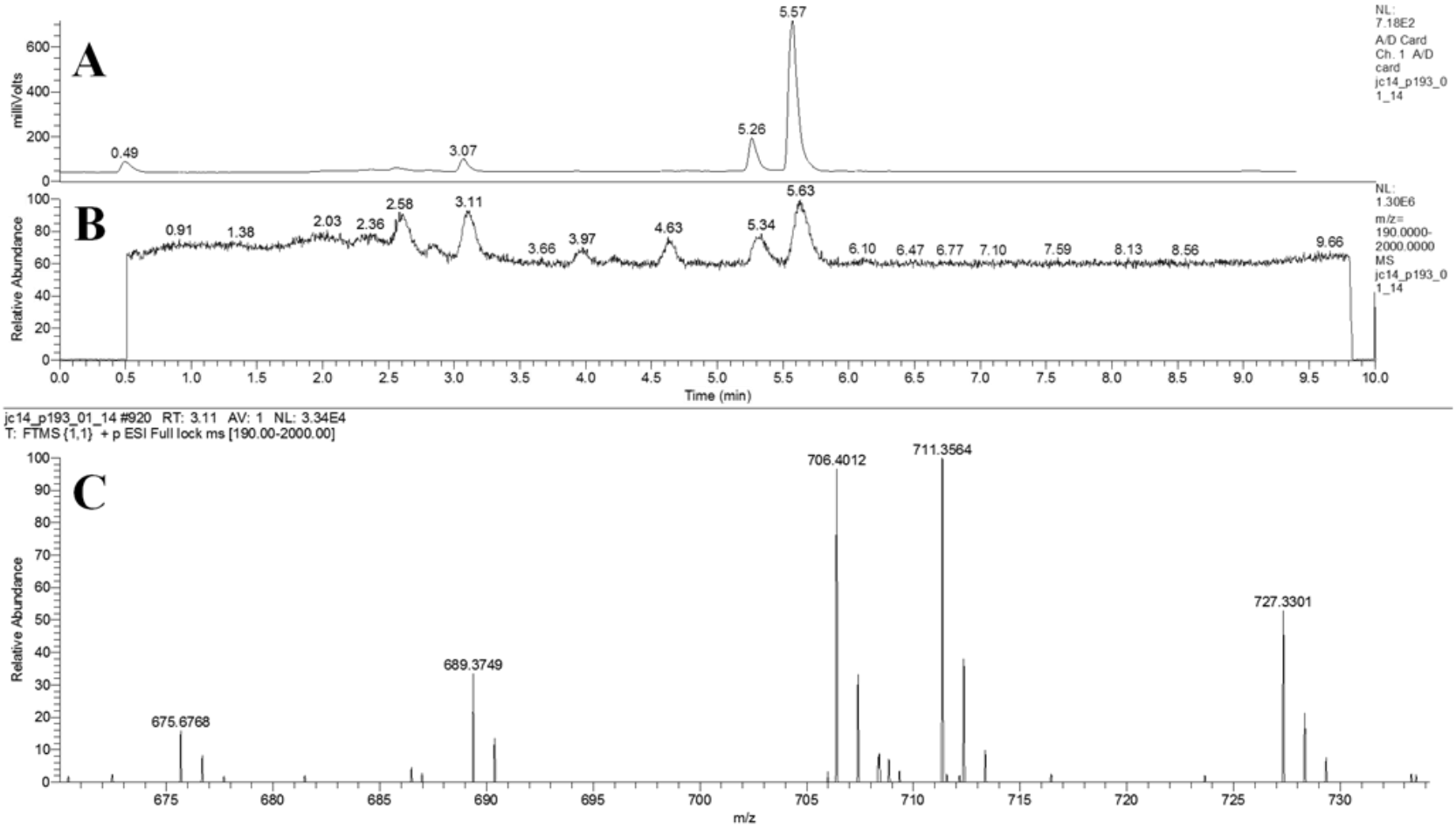


Figure 4.3. Evaporative light scattering detector (ELSD) **(A)**, mass chromatogram **(B)** and selected portion of the mass spectrum at 3.11 min **(C)** of the crude extract from *K. griseola* illustrating the production of satosporin A via the presence of the pseudomolecular ion $[M+H]^+$ and its sodium adduct *m*/*z* 689.3749 and 711.3564, respectively.

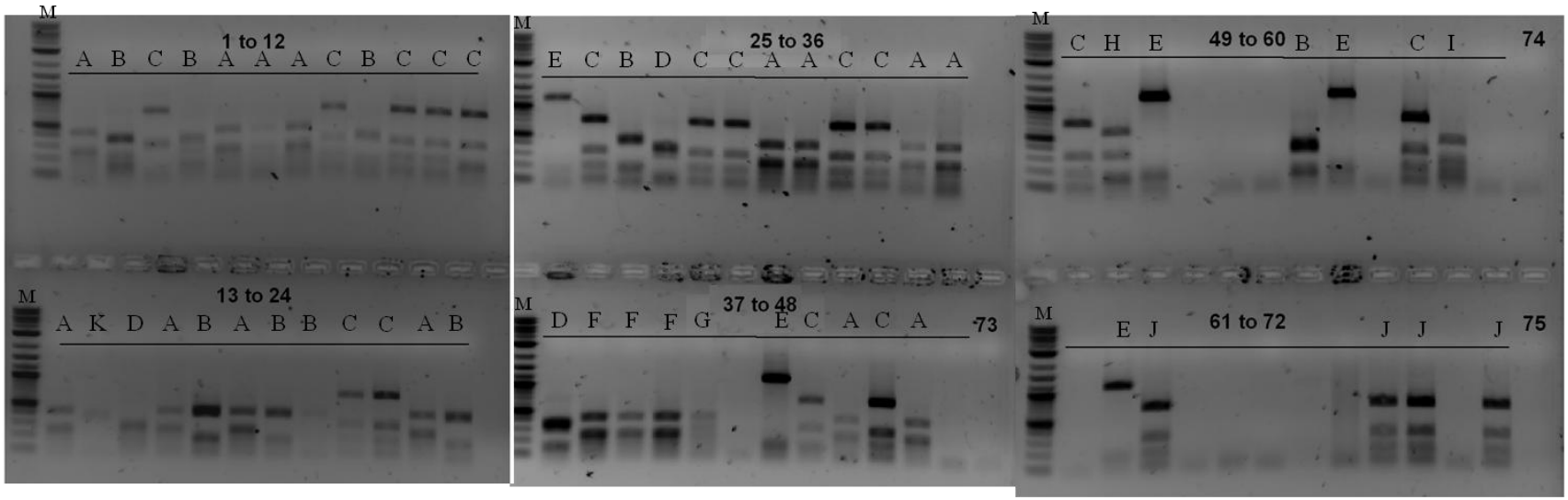


Figure 4.4. Restriction patterns, following digest and electrophoresis, of 75 KS-AT insert amplicons which grouped into 11 groups labelled from A to K; M: 2-log DNA ladder.

Table 4.2. Preliminary dereplication of each amplicon based on the 11 fragmentation patterns following XmaI and RsaI digest where the majority of amplicons have the restriction pattern of group A, B or C.

Group	Colonies	Group	Colonies
A	1, 5, 6, 7, 13, 16, 18, 23, 31, 32, 35, 36, 45, 47	G	41
B	2, 4, 9, 17, 19, 24, 27, 55	H	50
C	3, 8, 10, 11, 12, 21, 22, 26, 29, 30, 33, 34, 44, 46, 49, 58	I	59
D	15, 20, 28, 37	J	63, 69, 70, 72
E	25, 43, 51, 56, 62	K	14
F	38, 39, 40		

Table 4.3. Contigs generated by assembling the nucleotide sequences generated from sequencing all 58 KS-AT clone amplicons; 50 sequences grouped into 6 contigs while 8 sequences did not group into any contigs.

Contig	Colony
1	3, 8, 10, 11, 21, 26, 29, 30, 33, 34, 46, 49, 58, 70, 72
2A	2, 12, 50
2B	4, 9, 37, 55
3	1, 6, 7, 13, 15, 16, 18, 23, 32, 35, 36, 38, 40, 41, 45, 47, 59,
4	43, 56
5	17, 19, 20, 24, 27
No contig	14, 22, 25, 28, 39, 51, 62, 69

To further support the sequence-based groupings and to provide information as to which biosynthetic gene cluster these sequences originated from, a nucleotide BLAST® search was performed using a local database containing the bafilomycin PKS nucleotide sequences from *K. griseola* generated during this research. Of the 54 amplicons sequenced, 49 had greater than 94% sequence identity to a ketosynthase and/or methylmalonyl-CoA transferase region of the bafilomycin gene cluster, as listed in Table 4.4. The groups generated using the BLAST®n results were very similar to the contigs except that better resolution was provided for sequences that had not grouped into a contig. The sequence of colony 22 was now clearly part of group I as it shared 99% nucleotide identity to the same KS-AT region of bafilomycin module 1 and those of colonies 25, 51 and 62 could now be grouped in IV as they shared 99% sequence identity with the acyltransferase of module 7. The analysis also revealed that although colonies 69, 39, 28 and 14 did not have duplicates, these KS-AT regions were those of modules 1, 5, 8 and 10, respectively of the bafilomycin gene cluster.

In the case of KS-AT amplicons from modules 1, 4 and 5 (groups I, II and III), two contigs and/or groups were generated for each amplicon representing the two insertional directions illustrated in Figure 4.5 that are possible during cloning of the PCR products. Since the cloning strategy took advantage of the adenosine overhang produced by the Taq polymerase; the PCR products could thus be cloned into the plasmid in either direction[135].

According to the published bafilomycin gene cluster, the modules containing a methymalonyl-CoA transferase domain, and thus the sequence recognized by the degenerate primer M6R, were those of modules 1, 3, 4, 7, 8, 9, 10 and 11. Although the acyltransferase of module 5 had substrate specificity for methoxymalonyl-CoA, the M6R primer was successful in annealing to the DNA as evidenced by the amplicons of group III. Further inspection of the sequence of this domain revealed that the sequence used to design the primer was conserved in this module loading methoxymanolyl-CoA as its sequence differed by only one base at position 11 of the primer. We were thus successful at obtaining amplicons for two thirds of the possible

Table 4.4. Groups of KS-AT clones and associated modules of the bafilomycin biosynthetic gene cluster of *K. griseola*.

Group[a]	Colony	Bafilomycin domains in *K. griseola*[b] (nucleotide sequence identity>94%)
I-F	3, 8, 10, 11, 21, 22, 26, 29, 30, 33, 34, 46, 49, 58, 70, 72	KS_1→AT_1 *(BfaS1-AT_2)*
I-R	69	*AT_1 (BfaS1-AT_2)*
II-F	2, 12, 50,	KS_4→AT_4 *(BfaS2-AT_1)*
II-R	4, 9, 37, 55	*AT_4 (BfaS2-AT_1)*
III-F	39	KS_5→*AT_5 (BfaS2-AT_2)*
III-R	1, 6, 7, 13, 15, 16, 18, 23, 32, 35, 36, 38, 40, 41, 45, 47, 59	*AT_5 (BfaS2-AT_2)*
IV	25, 43, 51, 56, 62	*AT_7 (BafS3-AT_1)*
V	28	*AT_8 (BafS3-AT_2)*
VI	14	*AT_{10} (BafS4-AT_2)*
other	17, 19, 20, 24, 27	No bafilomycin similarity

[a]F for a forward and R for a reverse insertion of the ketosynthase-acyltransferase (KS-AT) amplicon in *E. coli*; [b]domain numbering was based on the overall extender module number while numbering in parenthesis refers to the synthase and the module within that synthase.

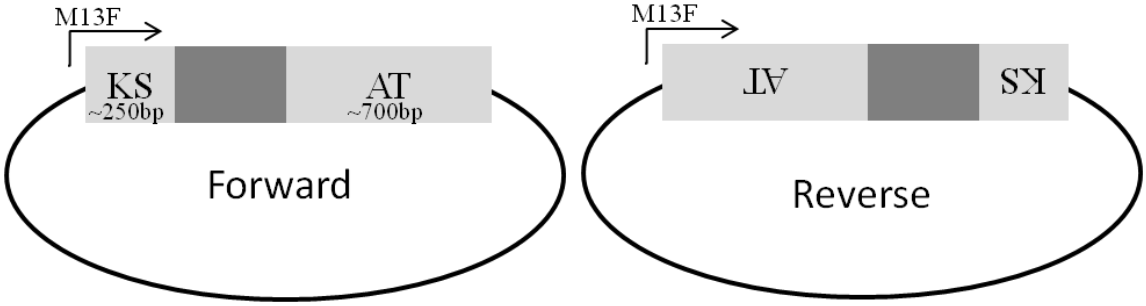


Figure 4.5. Graphical representation of the possible orientations of the KS-AT amplicon during cloning that explains why two groups were generated for amplicons sharing high sequence identity to the same bafilomycin domains.

bafilomycin KS-AT regions; increasing the number of clones sequenced could have generated better coverage. However, given that the objective was to obtain KS-AT sequences that did not belong to the bafilomycin cluster, further cloning and sequencing was not necessary.

The sequences of colonies in contig 5 were the only ones without high nucleotide identity to bafilomycin genes, suggesting that these amplicons did not originate from the bafilomycin cluster and, thus, could potentially represent a PKS KS-AT region from the satosporin biosynthetic gene cluster. It was also possibly amplified from a KS-AT region of another polyketide gene cluster as actinomycetes are known to have numerous secondary metabolite biosynthetic gene clusters in their genomes[30, 136]. When the sequences from amplicons in contig 5 were submitted to BLAST®x against the NCBI protein database, most hits had only 65% sequence identity to PKS genes, suggesting that these proteins shared similar functions but did not belong to the same gene clusters[137].

Illumina® sequencing

To obtain the DNA sequence of the entire satosporin biosynthetic gene cluster, Illumina® sequencing was performed on *K. griseola* genomic DNA. The sequencer generated 3,048,153 reads (150 bp), which first grouped into 3291 nodes between 26.427 kb to 61 bp long with a median length of 1192 bp and where half of the assembly could be generated using fragment lengths of 4467 bp or longer (N50). These nodes assembled into 596 contigs ranging between 143 and 32,727 bp in length with an N50 contig size of 8942 bp. Of the 534 contigs (>1 kb) and 390 nodes (>5 kb) submitted to antiSMASH, 30 sequences contained secondary metabolite genes and are listed in Table 4.5. Thirteen were assigned to portions of the bafilomycin gene cluster by performing a local BLAST®x search against protein sequences from the bafilomycin gene cluster in *K. setae*, while one sequence, node 345, had high sequence identity to the late portion of the terpentecin biosynthetic gene cluster in the same strain of *K. griseola*. The remainder of the sequences encoding for polyketide biosynthetic genes were analysed further by submitting them to BLAST®x against the NCBI database.

Table 4.5. Secondary metabolite biosynthetic genes in nodes and contigs generated from sequencing of *K. griseola* gDNA using Illumina® technology as detected by antiSMASH.

Sequence	antiSMASH hit[a]	Domains detected[b]	Originating Gene cluster	Length (bp)
contig 1	Type I PKS	AT-DH-ER-KR-ACP-KS		6995
contig 222	Type I PKS	KS-AT-DH		3651
node 439	Type I PKS	KS-AT		3645
node 697	Type II PKS	KS		10593
contig 95	Type I PKS	KS-AT-DH-ACP		23531
node 2086	NRPS-PKS			6143
contig 141	NRPS-PKS	KS-AT-DH-KR-ACP-C-A-T-TE	Possible homolog of frontalamide	16424
contig 277	NRPS-PKS		KSE_27190	18466
contig 278	Lantibiotic			17679
node 840	Squalene/hopanoid			8465
contig 147	Squalene/hopanoid			13186
contig 144	Siderophore			14561
node 392	Siderophore			10524
contig 56	Other			32727
contig 45	Other		KSE_61120	18541
contig 338	Other			7788
contig 4	Type I PKS		Bafilomycin	22436
contig 98	Type I PKS		Bafilomycin	15766
contig 60	Type I PKS		Bafilomycin	11823
contig 369	Type I PKS		Bafilomycin	6749
contig 78	Type I PKS		Bafilomycin	4195
contig 553	Type I PKS		Bafilomycin	1409
node 3541	Type I PKS		Bafilomycin	22436
node 1072	Type I PKS		Bafilomycin	15766
node 3611	Type I PKS		Bafilomycin	11823
node 253	Type I PKS		Bafilomycin	6749
node 127	Type I PKS		Bafilomycin	3330
node 598	Type I PKS		Bafilomycin	3282
node 3223	Type I PKS		Bafilomycin	1409
node 345	Terpene		Terpentecin	21158

[a]PKS: polyketide synthase, NRPS: non-ribosomal peptide synthetase, [b]AT: acyltransferase, KS: ketosynthase, ACP: acyl carrier protein, KR: ketoreductase, DH: dehydratase domains.

Sequences from contigs 141 and 277 as well as node 2086 were ruled as unlikely satosporin gene cluster fragments as no NRPS genes were predicted to be necessary during biosynthesis. This left contigs 1, 95 and 222 as well as nodes 439 and 697 as possible satosporin fragments as they all encoded for PKS domains that would be consistent with the predicted domains required for biosynthesis; contig 1 could represent the sequence of modules 2 or 6 whereas that of contig 222 may be that of modules 2, 5, 6, or 7 of the predicted pathway shown in Figure 4.2. The modules predicted from nodes 439 and 697 could belong to any modules as they contained the minimal domains required for chain elongation during polyketide biosynthesis; no module could be correlated with the domains predicted from contig 95 as DNA coding for the ketoreductase (KR) domain seems to be lacking and thus the dehydratase predicted here would have no hydroxyl group to act upon.

Since a KS-AT region had been amplified from degenerate primers (Section 4.3.2) that did not belong to the bafilomycin biosynthetic gene cluster, a locat BLAST®n was performed with sequences from colonies 17, 19, 20, 24 and 27 against the 3391 nodes generated from Illumina® sequencing. All five sequences had 94-100% sequence identity to region 669-1 of the 2962 bp region of node 2491. Although this region had not been identified by antiSMASH as secondary metabolite biosynthetic genes, a BLAST®x analysis revealed the presence of a partial acyltransferase, ketoreductase, acyl carrier protein and partial ketosynthase domains. This region would be consistent with modules 1 and 3 of the predicted satosporin biosynthetic gene cluster; it could however also belong to another PKS that produced a hydroxyl group during polyketide backbone assembly.

PacBio sequencing

Since Illumina® sequences failed to provide a contig containing the entire satosporin biosynthetic gene cluster, *K. griseola* gDNA was sent to the McGill University and Génome Québec Innovation Centre to be sequenced using PacBio technology. This technology was chosen here for its ability to produce long raw reads (~2 kb) to enable *de novo* assembly due to

its single molecule real-time (SMRT®) cell technology. Similar to Illumina® technology, PacBio sequencing provides sequencing by synthesis without the need to generate PCR amplified clusters of sequences. Instead, a single DNA molecule is bound to a DNA polymerase and immobilized in a well, called zero-mode waveguides (ZMWs), which then emits different light wavelengths as different bases are incorporated and their fluorophore cleaved[138]. Although the sequencing accuracy of a single read is poor (85-87%), resequencing allows for 99.09% accuracy after 7 passes[139]. The high coverage (>100X) obtained during this experiment was sufficient to overcome the low accuracy of single reads to generate high accuracy sequences.

The assembly contained 15 contigs (quiver_34-48) ranging in size between 2.8 kb and 2.3 megabase pair (Mb); more details on these data will be provided in Chapter 5 of this thesis. A local BLAST®n of non-bafilomycin KS-AT clone sequences (17, 19, 20, 24 and 27) and potential sequences highlighted by Illumina® sequences (contigs 1, 95 and 222 and nodes 439 and 697) against the assembled contigs generated from PacBio sequencing revealed which gene clusters had been previously detected in the genome of *K. griseola* as summarized in Table 4.6. Contig 222 and node 697 had high sequence identities with quiver_36 and when analysed by antiSMASH, revealed a NRPS-PKS gene cluster similar to that found in *Streptomyces clavuligerus* and *Streptomyces griseus*; the domain organization was clearly inconsistent with the predicted satosporin biosynthetic gene cluster. The region in quiver_42 which aligned with the sequence of contig 95 contained a single module with domains KS-AT-DH-ACP; the other necessary domains were absent from the surrounding regions and thus there was no evidence to suggest that this was the satosporin biosynthetic gene cluster.

Contig 1, node 439, node 2491 and the sequences of the non-bafilomycin KS-AT amplicons were revealed to be sequences present within quiver_44. An antiSMASH analysis of this region revealed a region between 467’063 and 566’778 containing, in addition to other genes, two large PKS containing 5 and 3 modules which was consistent with the one loading and seven extender modules predicted for satosporin biosynthesis. Contig 1 and node 2491 were both

Table 4.6. Alignment results between the potential satosporin gene cluster fragments discovered using degenerate primers and Illumina® sequencing with the 15 contigs (Quiver_xx) generated using PacBio genome sequencing technology.

Hits	Alignment hit	Alignment region[a]	% identity
Contig 222	Quiver_36	Q: 276-3651 S: 1'785'414-1'788'768	100% with misassembly
Node 697	Quiver_36	Q: 1-10'593 S: 1'789'159-1'799'751	99%
Contig 95	Quiver_42	Q: 2182-22'907 S: 1'690'212-1'669'487	99%
Contig 1	Quiver_44	Q: 552-4266 S:515'829-511'958 and other alignments	91-100% with misassembly
Node 439	Quiver_44	Q: 1-3645 S: 493'004-489'360	100%
Node 2491	Quiver_44	Q: 1-2992 S: 542'212-539'221	100%
KS-AT clones (17, 19, 20, 24, 27)	Quiver_44	Q: 1-732 S: 541'536-542'275	93-100%

[a]Q: query, S: subject.

sequences from these PKSs, while node 439 was present within another gene within the cluster. This suggested that the two PKSs in quiver_44 between 506'676 and 546'778 bp were probably responsible for the biosynthesis of the polyketide backbone of satosporin.

4.3.3 Insertional inactivation of module 1 to generate *K. griseola* ∆sat

To assess the involvement of the probable gene cluster present within quiver_44, the apramycin resistance gene was inserted within the sequence of the PKSs' first extender module (module1, Figure 4.6 (A)).

To achieve the insertional inactivation of module 1, the deletion plasmid N2491KO was designed using PCR amplified 1.5 kb homologous arms and the apramycin resistance gene within the suicide plasmid pSET151. Once the pieces were ligated and *E. coli* was transformed with the ligation reaction, pDNA was restricted from colony 6, 7 and 10 which had shown the appropriate amplicon between 3 and 4 kb during colony PCR. Although the fragments generated from colony 7 following digestion with AatII enzyme were not consistent with the predicted fragment sizes, those of colonies 6 and 10 were consistent in both restriction digests and thus pDNA from *E. coli*::pN2491KO colony 6 was used for the transformation of *E. coli* ET12567 harbouring pUZ8002 with the deletion plasmid, followed by conjugation with *K. griseola* MF730-N6.

Genetic characterization

Exconjugants were subcultured three times on media containing apramycin and nalidixic acid and then on media with apramycin and media with thiostrepton to screen for double cross-over exconjugants. Genomic DNA from clones that seemed resistant to apramycin but sensitive to thiostrepton was used to characterize the genotype using primers that annealed within the homologous arms used for the deletion plasmid assembly. A 1538 bp amplicon was expected for wild-type and single cross-over colonies, while a single 2493 bp amplicon was expected for double cross-over colonies. Clones 1, 2 and 4 showed amplicons consistent with a double cross-over of homologous arms as evidenced by a single band between 2 and 3 kb following gel

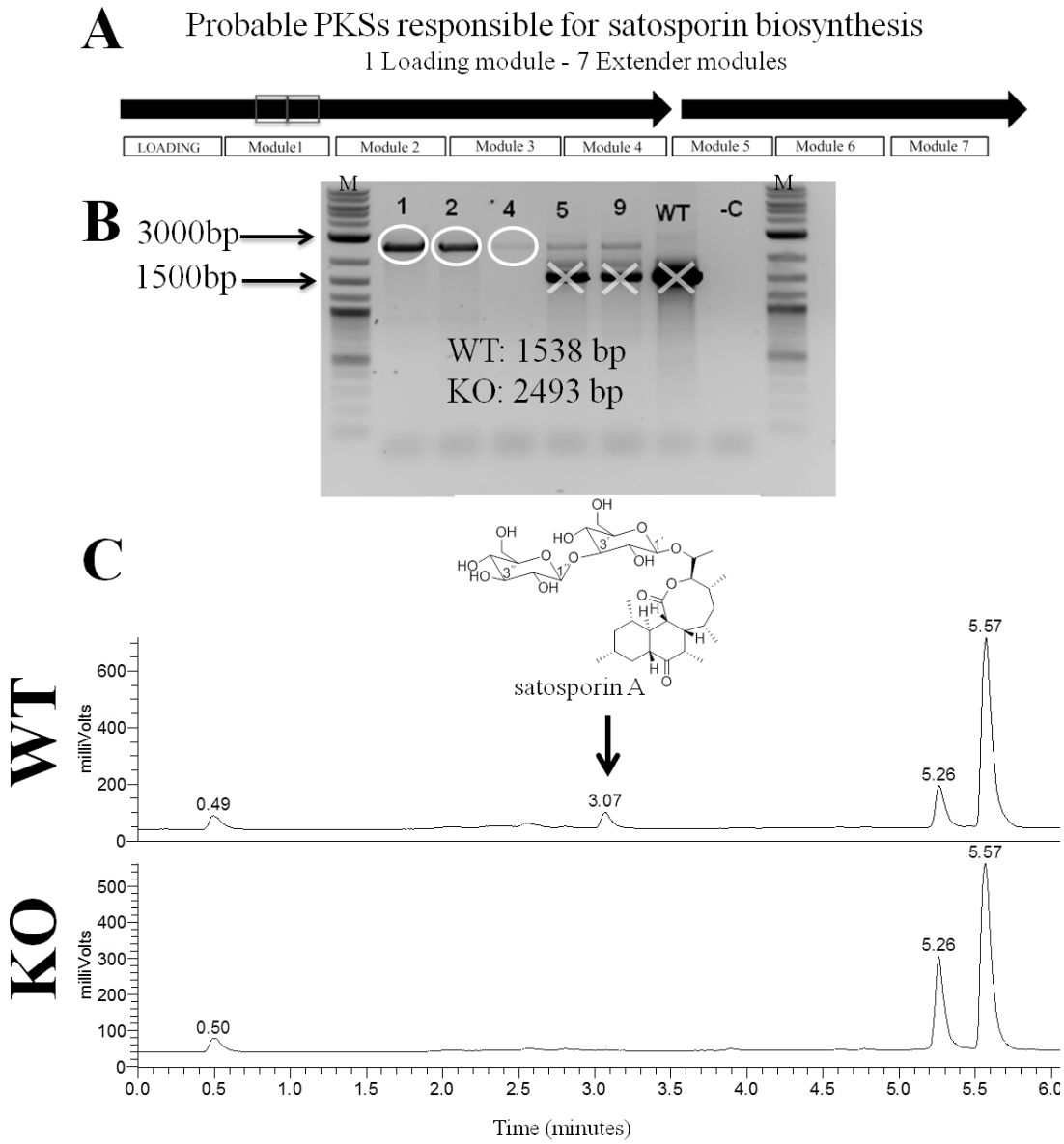


Figure 4.6. Probable satosporin biosynthetic genes present in quiver_44; regions of homology used for the insertional inactivation procedure are boxed in grey **(A)**, amplicons generated from genomic DNA of apramycin resistant and thiostrepton sensitive knock-out (KO) clones as well as from gDNA from the wild-type (WT) as a positive control and a negative control (-C) was also included where no DNA was added to the PCR reaction; bands consistent with a double cross-over of homologous arms are circled while those consistent with the native genotype are crossed-out, M: 2-log DNA ladder **(B)** and evaporative light scattering detector (ELSD) chromatograms of wild-type (WT) strain and knock-out (KO) strain; satosporin A was detected at 3.07 min in wild-type strains and was not detected in crude extracts of the knock-out strain where the apramycin resistance gene replaced 38 bp of sequence in module 1 of the probable satosporin biosynthetic PKS gene **(C)**.

electrophoresis (Figure 4.6 (B)). The wild-type strain showed an amplicon of approximately 1.5 kb, consistent with the native genotype and, finally, clones 5 and 9 showed amplicons of both lengths, representing most likely a single cross-over event even though the colonies seemed sensitive to thiostrepton. It is possible that these clones simply grew slower than their counterparts plated on media containing apramycin; verifying the genotype of the strains was thus a crucial step to choose a correct clone for phenotypic characterization. Since exconjugant clone # 1 showed genotype consistent with the desired insertional mutation, it was renamed *K. griseola* Δsat.

Phenotypic characterization

The phenotype of *K. griseola* Δsat was then compared to that of the wild-type following 48-hour fermentations in lean medium and EtOAc extraction. From the ELSD chromatograms (Figure 4.6 (C)), satosporin was detected in the crude extract of the wild-type strain, as evidenced by a peak at 3.07 min with *m*/*z* of 711.3575 in the mass spectrum. The crude extract from fermentations of the knock-out strain failed to show an ELSD peak around 3 min and no ions with *m*/*z* 711.3562±5 ppm could be detected in any mass spectrum. There was no evidence of analogues or shunt metabolites.

Although a complementation study was not performed here, there is evidence to suggest that the phenotype was not a result of polar effects causing failed transcription in genes at other loci on the genome. Firstly, there were no apparent differences in growth rates or sporulation between knock-out and wild-type strains and secondly, the knock-out strain was still capable of producing other secondary metabolites as evidenced by the peaks at 5.26 and 5.57 min which had been shown to be bafilomycins.

The evidence supports the hypothesis that the polyketide genes present between base pairs 506’676 and 546’778 of quiver_44 encode for the two PKSs responsible for biosynthesizing the backbone structure of satosporins.

4.3.4 Sequence analysis of the *sat* cluster and revised proposed biosynthesis

Sequence analysis

AntiSMASH predicted a gene cluster of 99.7 kb containing the PKSs involved in satosporin biosynthesis; it spanned nucleotides 467'063 to 566'778 of quiver_44 and had a G+C content of 73.5%, consistent with that of *Kitasatospora* reported previously[136].

ORF10 and *ORF1* were proposed as the boundaries of the sat cluster, spanning the 67.590 kb region between bases 481'188 and 548'778 of quiver_44. Given that most genes were transcribed from the complimentary strand (from right to left), the DNA sequence of quiver_44 was reverse-complemented to discuss the sat cluster as illustrated in Figure 4.7. The end of the cluster was predicted to be *ORF10* due to the presence of genes with homology to transposases (*ORF11-12*) downstream of the cluster. Since horizontal gene transfer is considered to play an important role in gene cluster formation, mobile genetic elements such as transposons and DNA repeats commonly flank secondary metabolite biosynthetic gene clusters[140]. Since no role in satosporin biosynthesis could be attributed to genes upstream of *ORF1*, the beginning of the cluster was proposed to be with *ORF1*. Knock-out experiments would be necessary to confirm that the genes outside the delimited cluster were not involved in satosporin biosynthesis.

The open reading frame prediction was confirmed using FramePlot and the resulting protein sequence was used to query the NCBI protein database using BLAST®p; hits for similar proteins to each ORF are listed in Table 4.7 where identity between query and database sequences varied between 33 and 90% whereas their similarity varied between 50 and 94%. Given the high sequence similarities (>40%) between ORFs and the database, it was possible to extrapolate and propose functions, as listed in Table 4.8, for most of the newly sequenced proteins within the *sat* cluster of *K. griseola*[141].

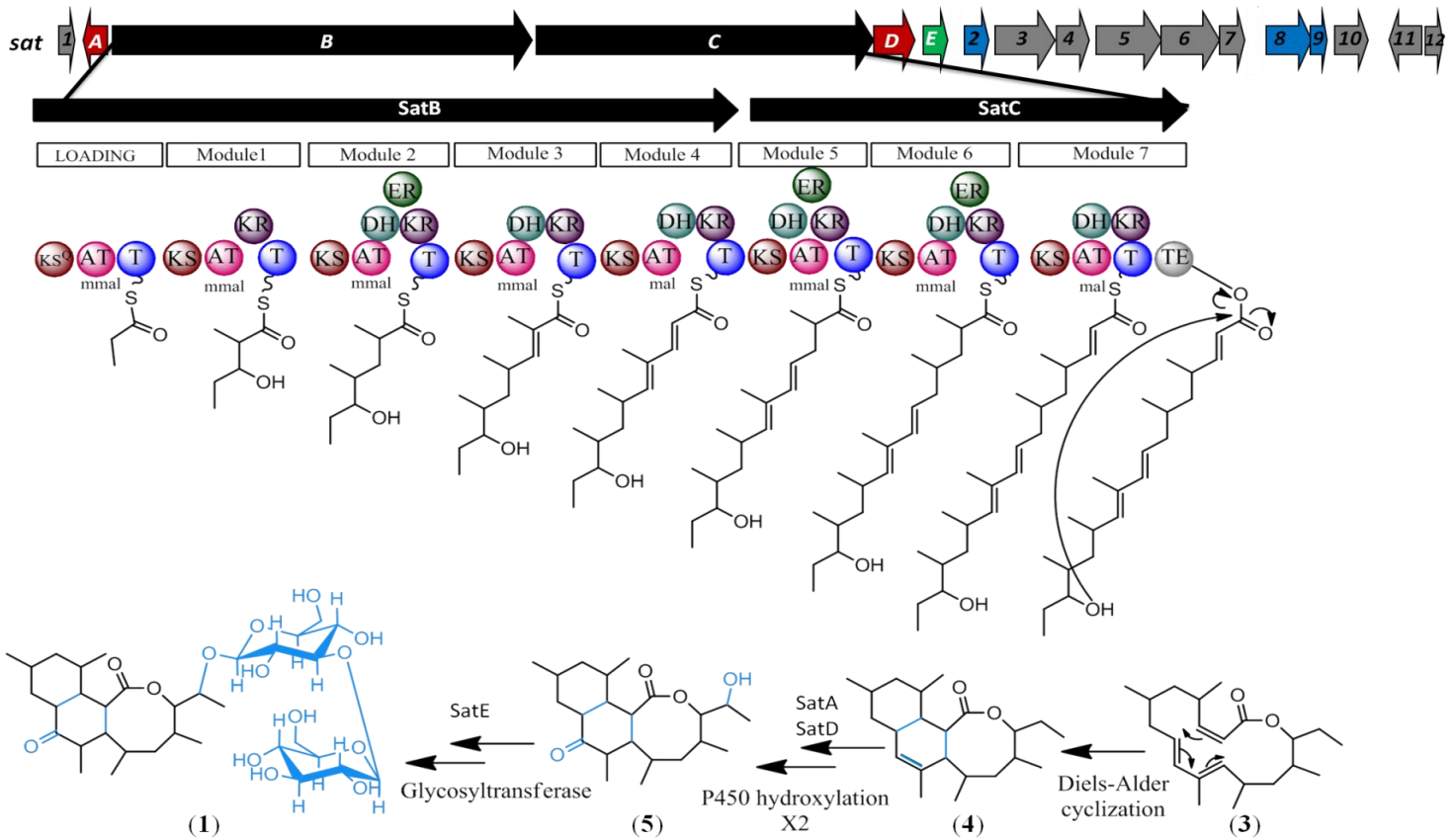


Figure 4.7. The proposed biosynthesis of satosporins based on organization of the gene cluster involves two large multimodular polyketide synthase genes (*satB and satC*, black arrows) as well as tailoring enzyme genes encoding P450 monoxygenases (*satA* and *satD*, red arrows) and a glycosyltransferase (*satE*, green arrow). Regulatory and self-resistance genes (*ORF2*, *8* and *9*, blue arrows) were also found within the *sat* cluster in *K. griseola*; genes not predicted to be involved are shown in dark grey and stereochemistry is not shown to simplify visualization of the Diels-Alder intramolecular cyclization, mmal:methylmalonyl-CoA and mal: malonyl-CoA extender units, AT: acyltransferase, KS: ketosynthase, ACP: acyl carrier protein, KR: ketoreductase, DH: dehydratase, ER: enoylreductase, TE: thioesterase.

Table 4.7. BLAST®p results for each open reading frame (ORF) within the satosporin biosynthethic gene cluser; aa, amino acid.

Protein	Size (# aa)	Similar protein, origin, (%identity/similarity), accession number
ORF1	298	Hypothetical protein, *Streptomyces roseochromogenes*, (66/79), WP_023545500.1
SatA	389	Cytochrome P450, *Amycolatopsis mediterranei*, (50/66), WP_013227439.1
SatB	7590	Polyketide synthase ObsC, *Saccharopolyspora spinosa,* (51/61), AAS00421.1
SatC	5774	Polyketide synthase Pks12, *Mycobacterium tuberculosis* M1283, (52/64), KAY30031.1
SatD	750	Beta-ketoacyl synthase-like protein, *Saccharopolyspora spinosa*, (53/69), WP_010696351.1, (aa 1-323) Cytochrome P450, *Nocardia brasiliensis*, (61/74), WP_029895429.1, (aa 349-748)
SatE	387	Antibiotic resistance macrolide glycosyltransferase, *Saccharopolyspora spinosa,* (52/63), WP_010692853.1
ORF2	429	MFS transporter, *Stackebrandtia nassauensis,* (33/53), WP_013020978.1
ORF3	1066	Amino acid adenylation domain protein, *Microcoleus vaginatus,* (33/50), WP_006635040.1
ORF4	765	Polyketide synthase family protein, *Oscillatoria acuminata,* (44/61), WP_015149440.1
ORF5	1145	Amino acid adenylation protein, *Streptomyces flavogriseus,* (39/50), WP_014155552.1
ORF6	953	Polyketide synthase, *Myxococcus xanthus,* (45/56), WP_011553648.1
ORF7	536	Peptide synthase (partial), *Pseudomonas* sp. Lz4W, (47/63), WP_003444907.1
ORF8	679	SARP family transcriptional regulator, *Catenulispora acidiphila,* (38/51), WP_015793868.1
ORF9	332	Putative regulatory protein, *Streptomyces atroolivaceus,* (71/80), AF484556_16
ORF10	510	Hypothetical protein KCH_77500, *Kitasatospora cheerisanenses* KCTC 2395, (90/94), KDN80516.1
ORF11	524	Transposase IS4 family, *Streptomyces ipomoeae*, (77/94), WP_009338417.1
ORF12	316	Transposase, *Streptomyces canus,* (77/84), WP_028808298.1

Table 4.8. Deduced function of each open reading frame (ORF) based on sequence similarity to previously annotated genes; AT: acyltransferase, KS: ketosynthase, ACP: acyl carrier protein, KR: ketoreductase, DH: dehydratase, ER: enoylreductase, TE: thioesterase, C: condensation, A: adenylation, PCP: peptidyl carrier protein domains.

Protein	Proposed function in satosporin biosynthesis and domains present when appropriate
ORF1	Unknown
SatA	P450 hydroxylation
SatB	Polyketide backbone assembly by type I PKS KS^{Q}, AT^{L}, ACP^{L}, KS^{1}, AT^{1}, KR^{1}, ACP^{1}, KS^{2}, AT^{2}, DH^{2}, ER^{2}, KR^{2}, ACP^{2}, KS^{3}, AT^{3}, DH^{3}, KR^{3}, ACP^{3}, KS^{4}, AT^{4}, DH^{4}, KR^{4}, ACP^{4}
SatC	Polyketide backbone assembly by type I PKS KS^{5}, AT^{5}, DH^{5}, ER^{5}, KR^{5}, ACP^{5}, $KS^{6,}$ AT^{6}, $DH^{6,}$ ER^{6}, KR^{6}, ACP^{6}, KS^{7}, AT^{7}, DH^{7}, KR^{7}, ACP^{7}, TE^{7}
SatD	P450 oxidation
SatE	Glycosylation of the aglycone
ORF2	Secretion/resistance
ORF3	Unknown C, A, PCP
ORF4	Unknown KS
ORF5	Unknown C, A, PCP
ORF6	Unknown KS, AT, ACP
ORF7	Unknown A
ORF8	Regulation
ORF9	Regulation
ORF10	Unknown
ORF11	Horizontal gene transfer
ORF12	Horizontal gene transfer

Polyketide backbone assembly by multimodular polyketide synthases SatB and SatC

Two large multimodular PKSs were found to be encoded by *satB* and *satC;* together they provided the loading module as well as the seven extender modules necessary to extend the loading unit by 14 carbon units.

SatB contained, at its N-terminus, a loading module with domains KS^Q, AT^L and ACP^L. Both the KS and AT domains had protein characteristics supporting an acylation and decarboxylation mechanism for polyketide chain initiation observed in many modular PKSs including those involved in pikromycin, oleandromycin and tylosin biosynthesis[9]. Firstly, the ketosynthase was structurally similar to other KSs; however, the cysteine (C) residue within the active site VDTACSSS was replaced by glutamate (Q)[142], a characteristic feature in all KS^Q, allowing the enzyme to decarboxylate the malonyl or methylmalonyl unit bound to ACP[9]. Secondly, the AT domain contained the conserved arginine[117] residue, which is present in loading and extender AT domains that interact with CoA esters of dicarboxylic acids such as malonyl- and methylmalonyl-CoA. This is in contrast to the loading AT domains that incorporate monocarboxylic acid-CoA such as acetyl- and propionyl-CoA, such as in the erythromycin AT^L for example, where the arginine residue is replaced by a tryptophan residue at this location[142]. The satosporin acyltransferase AT^L, had a predicted substrate specificity for methylmalonyl-CoA, as determined by the motif QQGHSQRSHANV, mostly consistent with the QQGHS[QMI]GRSHT[NS]V motif predicted for acyltransferases incorporating methylmalonate units[143]; the only difference was that the threonine (T) had been replaced by an alanine (A); a substitution observed in many acyltransferases of spirangiene PKSs[144]. These data suggest that satosporin biosynthesis begins with the loading of a methymalonyl unit onto the ACP phosphopantethienyl arm via a thioester bond followed by its decarboxylation by KS^Q to give rise to an ACP bound propionyl starter unit.

The seven β-ketoacyl-ACP synthases (KS) within the *sat* cluster, responsible for seven chain elongations via decarboxylative condensations shared on average 80% amino acid

sequence identity; KSs of modules 4 and 7 were most similar, sharing 89% sequence identity. Alignment with other KS domains of PKSs within the SBSPKS database, satosporin KSs consistently showed highest sequence identities of 74 and 73% with KS domains from either avermectin or amphotericin PKSs, respectively[134].

Of the seven acyltransferase (AT) domains required to load and transfer the next extender units onto the module's ACP, five had a substrate specificity sequence motif consistent with loading of a methyl-malonate unit as mentioned previously: AT^1, AT^2, AT^3, AT^5 and AT^6. In contrast, the sequence motif QQGHSVGRFHTQV of AT^4 and AT^7 was consistent with that predicted for acyltransferases incorporating a malonate unit: QQGHS[LVIFAM]GR[FP]H[ANTGEDS][NHQ]V[143]. Satosporin's backbone formation thus involves 5 (2S)-methylmalonyl-CoA and 2 malonyl-CoA extender units in the order illustrated in Figure 4.7.

All seven extender modules contained a β-ketoacyl-ACP reductase domain responsible for the stereospecific reduction of the β-carbonyl to a β-hydroxyl group using NADPH. All KR domains in satosporin biosynthesis appeared to be active as per the presence of the catalytic tyrosine (YAAAN motif) and NADPH binding site (TGGTGxLG motif) within each domain. Reducing KRs are divided into two main groups: B-type KRs contain an LDD motif ~57 amino acids (aa) before the catalytic tyrosine residue and result in hydroxyl groups with D-orientation, whereas A-type KRs do not contain the motif but instead have a tryptophan (W) 8 aa before the tyrosine residue and result in L-hydroxyl groups. Since the LDD motif and tryptophan residue are located on opposite sides of the active site, they are thought to guide the polyketide to enter at opposing sides of the active site, giving rise to the opposite stereochemistry following reduction. Domains KR^2, KR^3, KR^5 and KR^6 within the satosporin PKSs all had the LDD motif characteristic of B-type KRs whereas domains KR^4 and KR^7 had LAD and VDD motifs, respectively; these amino acid substitutions were previously reported for B-type KRs[145]. These KRs can be further subdivided into B2- and B1-type KRs, depending on whether or not the

enzyme carries out an epimerization reaction before reduction, resulting in L- and D-stereochemistry of the α-substituent, respectively. The presence of a proline 2 aa past the catalytic tyrosine residue would indicate B2-type KRs; however, KRs^{2-7} did not contain this residue, resulting in B1-type and thus D-stereochemistry for both the β-hydroxyl group and the α-substituent when appropriate. These observations were consistent with previous reports where B1-type KRs were usually involved in modules containing further processing enzymes[145, 146].

The only KR in satosporin PKSs whose product did not undergo further processing was KR^1; this enzyme did not contain the LDD motif but did have a tryptophan (W) 8 aa before the catalytic tyrosine, thus belonging to an A-type KR. Since no histidine was present 3 aa before the catalytic tyrosine, KR^1 was classified as an A1-type KR and would result in L-stereochemistry for both the β-hydroxyl group and the α-substituent since no epimerization of the α-substituent would occur previous to reduction initiated by the attack of the pro-4*S* hydride of NADPH to the carbonyl[145, 146]. A histidine residue was, however, present 4 aa before the catalytic tyrosine, it is possible that this residue may be capable of sterically selecting against the unepimerized polyketide and thus allowing only epimerized products to react with the tyrosine[145]. However, given the rectus (*R*) stereochemistry at carbons 14 and 15 in satosporins, L-configurations are required for both the α-substituent and the β-hydroxyl groups following reduction by KR^1 suggesting that the first ketoreductase in satosporin PKS was more likely of A1-type.

Modules 2 to 7 contained dehydratase domains to further reduce the β-hydroxyl group to a double bond between the α and β-carbons of the acyl-ACP. All appeared to be functionally active as they contained the motif HxxxGxxxxP containing the catalytic histidine. Although no sequence motif is capable of predicting whether the dehydratase produces the *cis*- or *trans*-alkene, *cis*- acting DH are usually preceded by A-type KR. Since KR^{2-7} are of B1-type, it is predicted that the catalytic aspartic acid will donate a hydrogen to the D-β-hydroxyl group and when the catalytic histidine extracts the L-α-hydrogen, resulting in *syn* elimination of water from

the β-carbon, giving rise to a *trans*-alkene[146]. Since modules 3, 4 and 7 did not contain further processing domains, the proposed satosporin polyketide backbone contained 3 *trans*-double bonds.

Modules 2, 5 and 6 contained active enoylreductase domains, evidenced by the presence of the NADPH binding site GGVGMA, resulting in complete reduction of the initial β-keto groups[146]. Enoylreductases achieve this reduction via the formation of the enolate intermediate following the NADPH pro-*4R* hydride attack of the β-carbon followed by the α-carbon accepting a proton from a tyrosine or lysine residue[146]. Given the presence of the tyrosine ~94 aa before the NADPH binding site, satosporin ER domains were predicted to result in L-α-substituents, which is consistent with the observed L- stereochemistry at carbons 12, 6 and 4.

The satosporin thioesterase domain was located at the C-terminus of the second PKS (SatC) and appeared to be active due to the presence of the histidine-aspartate-serine catalytic triad necessary for macrocyclization and polyketide release. The nucleophilic serine would presumably first attack the carbonyl of the thioester bound acyl-ACP, releasing it from the phosphopantetheinyl arm of ACP and resulting in an ester linkage to the serine residue of the thioesterase[146]. Secondly, the basic histidine residue would remove a hydrogen from the secondary alcohol created from the satosporin PKS module 1, which would then attack the ester carbonyl leading to macrocyclization and release from the serine residue of the PKS[146], giving rise to the macrolactone (**3**) shown in Figure 4.7.

Finally, interpolypeptide linker sequences, also known as docking domains, are short (~120 aa) but important sequences at the joining termini of two PKSs which allow crucial substrate channelling between the synthases[147]. The Dock_Dom_Anal interface of SBSPKS detected four helix bundle that are characteristic of the PKS docking domains, three from the C-terminus of SatB and one from the N-terminus of SatC. Based on the DEBS2-3 docking domain, where two interacting ionic pairs are responsible for joining the two synthases, D64-K92 and R73-D105, only the aspartic acid-lysine interacting ionic pair was detected (D64-K92) in the

SatB-C docking domain[134]. The presence of this docking domain linking the C-terminus of SatB and the N-terminus of SatC supports the poposed order of polyketide assembly illustrated in Figure 4.7.

Diels-Alder cyclization

Following polyketide backbone assembly by the PKSs, SatB and SatC, satosporin biosynthesis is proposed to then undergo intramolecular cyclization via a Diels-Alder [4+2] cycloaddition between the electrophilic dienophile generated by module 7 and the diene generated by modules 3 and 4. The reactivity of this reaction is favored by the electron withdrawing ester functional group that activates the dienophile to this cycloaddition (Figure 4.8 (A)). Many natural products such as terpenes, alkaloids, and polyketides have been proposed to include a biosynthetic Diels-Alder reaction[148]; examples include solanapyrones[149], nargenicin[150], betaenone B[151], lovastatin[152], macrophomic acid[153] and spinosyn A[154].

Although few Diels-Alderase enzymes have been described, there was no genetic evidence to suggest that any gene within the *sat* cluster would encode an enzyme responsible for catalyzing a Diels-Alder reaction; there were no sequence similarities to the possible Diels-Alderases lovastatin nonaketide synthase (LovB)[152], macrophomate synhtase (MPS)[153], spinosyn synthase (SpnF)[155] or solanopyrone synthase (Sol5)[149]. The spontaneous nature of this reaction was further supported by the increased reactivity provided by the macrocyclization by SatC thioesterase which could provide structural rigidity towards the s-*cis* diene necessary for Diels-Alder cyclization. Furthermore, the *trans*-decalin product was consistent with the endo transition state and product (Figure 4.8 (B)) that is often observed, thus supporting a spontaneous Diels-Alder reaction. Knock-out experiments of individual genes within the *sat* cluster may provide further insights as to whether one may act as a Diels-Alderase or not.

P450 oxidations

The two main sources of oxygen in polyketides are either from the acetate/propionate-derived extender units incorporated during assembly by PKSs or by enzymatic oxidation using

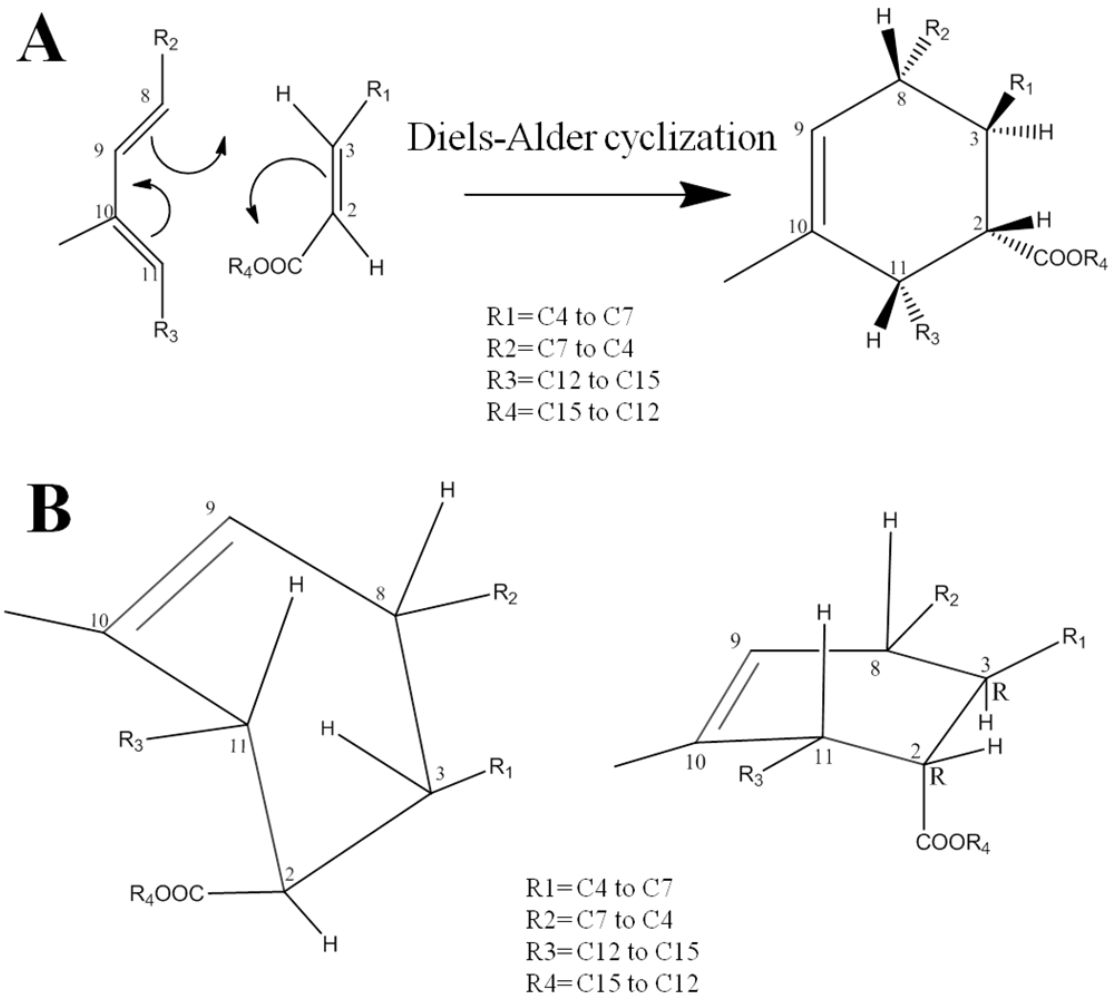


Figure 4.8. Diels-Alder cyclization reaction mechanism illustrating the *trans*-stereochemistry of the decalin moiety (rings fused at C3-C8) resulting from the s-*cis*-diene and *trans*-dienophile **(A)** and side views of the favoured endo product that led to the correct stereochemistry observed in satosporin **(B)**.

cytochrome P450 monooxygenases. Based on the domain organization within each module of the *sat* PKSs and the necessity for the diene and dienophile to allow the Diels-Alder intramolecular cyclization, the oxygen atoms at C9 and C16 were predicted to have arisen from P450 mediated oxidations.

Two putative cytochrome P450 monooxygenases were found within the *sat* cluster based on high amino acid sequence similarity to other CYP450. Sat A was found to share 66% sequence similarity with a cytochrome P450 from *Amycolatopsis mediterranei*, while the last 400 aa of SatD shared 74% sequence similarity to a cytochrome P450 from *Nocardia brasiliensis*. Cytochrome P450 monoxygenases carry out a variety of reactions including hydroxylations, epoxidations, *N*-oxidation, deamination, dehalogenation and *N*-, *O*- and *S*-dealkylation[156]. In order for the enzyme to oxidize the molecule, a heme cofactor must bind at a conserved cysteine residue while O_2 binds to a tyrosine; other components such as a flavin-dependent ferrodoxin reductase are also necessary to transfer electrons from NAD(P)H to heme. Both cytochrome P450 enzymes found within the *sat* cluster contained the conserved binding site residues tyrosine and cysteine as shown in the alignment of Figure 4.9, supporting their involvement in satosporin biosynthesis.

In the proposed biosynthesis of satosporins, hydroxylation is necessary at C16 to provide the proper functional group for glycosylation; one of the two P450s must thus perform this reaction. Given the typical CYP450 structure of SatA and the unusual structure of SatD, SatA is probably responsible for this typical alkane hydroxylation at C16. Similar macrolide tailoring hydroxylations are common in natural products and have been observed in compounds such as erythromycin[157], lankamycin[158], pikromycin[91], fostriecin[159], fusarin[160] and nargenicin[161].

The second oxidation at C9 is less straight forward as this location contains a C9-C10 alkene as a result of the Diels-Alder cyclization. Two mechanisms for oxidation are thus suggested to convert the alkene to the ketone present in satosporins. The first, involves a mechanism where, following formation of the olefin addition intermediate (**ii**), a rearrangement

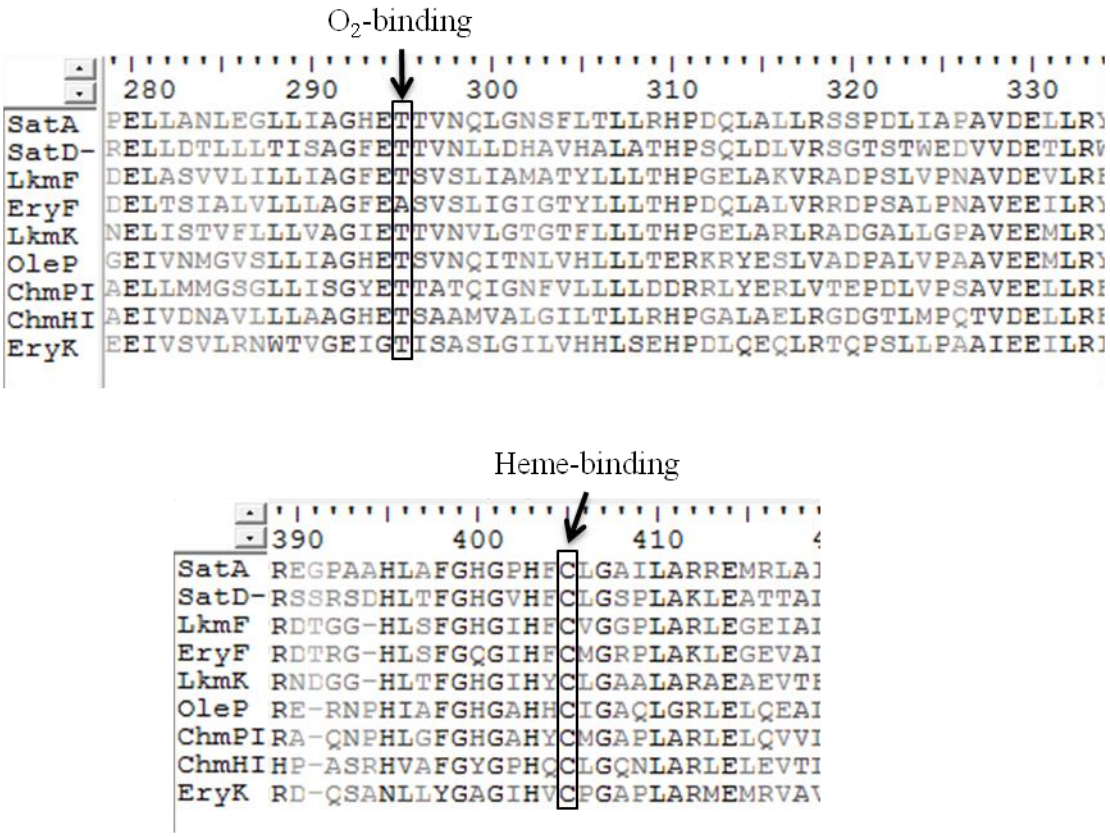


Figure 4.9. Partial alignment of the two cytochrome P450 sequences found in the sat cluster (satA and late portion of SatD) with other natural product P450s illustrating the presence of the necessary O_2 and heme binding sites to confer activity as oxidative enzymes; LkmF: lankomycin C8 hydroxylase, EryF: EM C6 hydroxylase, OleP: oleandomycin C8 hydroxylase, ChmPI: probable chalcomycin C8-hydroxylase, ChmHI: probable chalcomycin C20-hydroxylase, EryK: EM C12 hydroxylase; modified from Arakawa et al.[158]

of the electrons and hydride shift to yield the ketone (**iii**), rather than the attack of the carbonium or radical to form the epoxide[162, 163] (Figure 4.10). The second involves a mechanism where the P450 would hydroxylate the alkene at C9, forming the enol shown in Figure 4.11, which could then tautomerize to the ketone to provide structure (**5**) in Figure 4.7. Many P450 carryout unusual reactions such as ring expansion, isomerization via oxidation, halogen oxygenation, etc.[162] and, thus, an unusual role for this structurally unique protein is proposed.

Whichever mechanism is involved here, it is possible that betaenone B biosynthesis involves a similar pathway given that a ketone must also be formed from the octalin produced from a Diels-Alder cyclization (Figure 4.12). The origin of the ketone oxygen was proven to not arise from acetate or propionate via [1-^{13}C, $^{18}O_2$] labelling studies and was instead incorporated by P450 monoxygenase enzyme(s) since the non-oxygenated precursor, probetaenone I, was obtained when potent P450 inhibitors were included in the fermentation medium[151]. Unfortunately, no biosynthetic gene cluster has yet been described for betaenone.

Although some CYP450 are known to carryout multiple oxidation reactions at different locations, such as PikC in pikromycin/methymycin biosynthesis[91], most bacterial natural product P450s are highly specific and carryout a single reaction[164], such as EryF and EryK for C6 and C12 hydroxylation, respectively[157, 165]. Although it was not possible to conclude whether SatA and/or SatD are involved in oxidation at C16 and C9, it may be proposed that the unique features of SatD may be responsible for the unique transformation at C9.

SatD was unique in that it appeared to be a fused protein, where the first 323 aa showed sequence similarity to β-ketoacyl synthase III proteins, while the remaining 400 aa showed similarity to cytochrome P450 enzymes. No protein within the NCBI database showed similarity over the entire query, suggesting that such a fused KS/P450 protein has not been previously described. Although P450 enzymes have been reported as fusion proteins such as P450/ NADPH:p450 reductase[166] and dioxygenase/P450[167]. Sequencing of PCR amplicons of the aa region 240-452 revealed no errors in which a stop codon could have been overlooked.

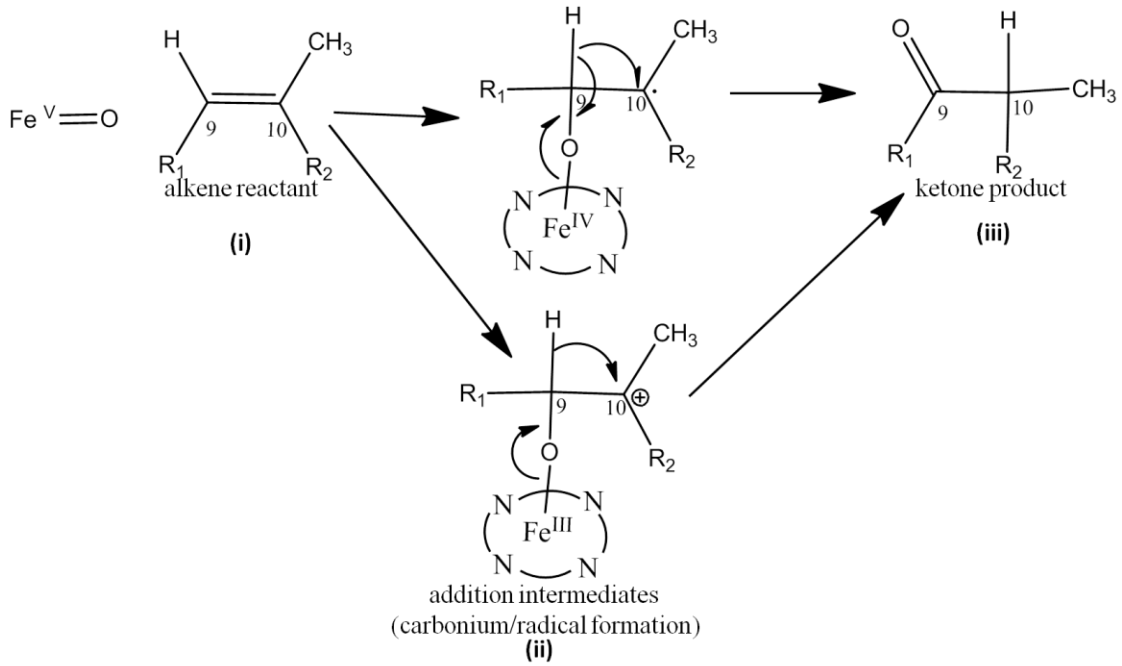


Figure 4.10. π-bond oxidation reactions by P450 monoxygenases, modified from Guengerich and Macdonald[163] to explain ketone formation from the alkene generated during satosporin biosynthesis which involves electron rearrangement and a hydride shift.

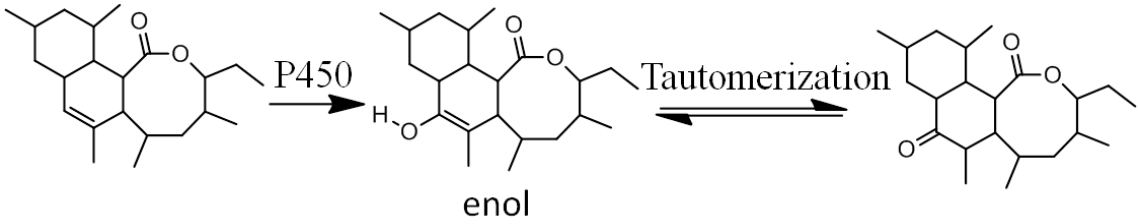


Figure 4.11. Second proposed mechanism to afford the satosporin ketone at C9, which involves hydroxylation of the alkene followed by tautomerization of the enol to the ketone.

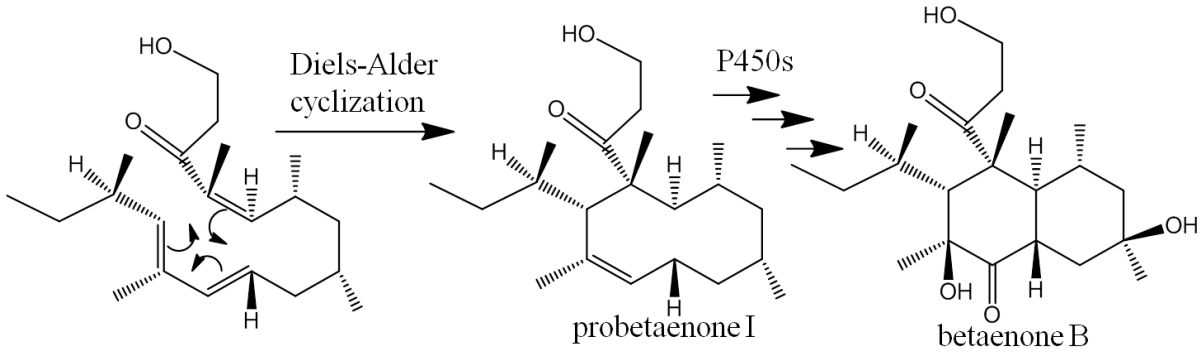


Figure 4.12. Betaenone B biosynthesis may also involves a P450 to mediate the oxidation of the alkene generated by a Diels-Alder cyclization to form a ketone; modified from Stocking and Williams[148].

Although the P450 domain had residues supporting an active role in oxygenation, the KS-like domain of the protein did not contain the conserved catalytic cysteine, but rather has a serine at this position instead. This has been observed in KS-like proteins such as CosE, DpsC, PksF, AknE2 and TaK, which are hypothesized to function either as fidelity factors towards the loading unit used by the PKS or involved in some tailoring reaction following PKS release of the polyketide [168, 169]. Further investigation of this unique enzyme would yield a better prediction for its role in satosporin biosynthesis.

To confirm the oxidative roles of SatA and SatD in satosporin biosynthesis, individual knock-out experiments or fermentations with potent P450 inhibitors are necessary to determine whether or not deoxygenated analogues would be produced; this may also reveal which P450 acts at which site and in which order. It may also provide insights as to whether these oxygenation reactions occur before or after the Diels-Alder cyclization. The oxidations at C9 and C16 would thus provide the satosporin aglycone (**5**), described previously as satosporin C[129].

Sequential glucosylation via a single glycosyltransferase

To convert satosporin C into the glycosylated natural products satosporin A and B, one and two glycosylation reactions must occur, respectively. The sat cluster contained an ORF, *satE,* with high protein sequence identity to macrolide glycosyltransferases; the closest match for SatE was the antibiotic resistance macrolide glycosyltransferase from *Saccharopolyspora spinosa*, sharing 63% sequence similarity.

Most glycosylated bacterial natural products contain a large diversity of deoxygenated sugar moieties that often play an important role in biological activity[170]. Surprisingly, satosporins incorporate an unmodified sugar moiety, glucose. Vancomycin[115] and pyralomycin 2C[117] are two examples of natural products that contain a glucose moiety and while the rebeccamycin and antibiotic AT2433 glycosyltransferases RebG and Atm incorporate a glucose molecule to the indocarbazole aglycone, it is then *O*-methylated to give the 4-*O*-methylglucose present in the natural products[171, 172].

Other natural products may use glucosylation as a self-defense mechanism. Such is the case in oleandomycin and methymycin/pikromycin, where glucosyltransferases yielded inactivated glucosylated natural products that required β-glucosidases to cleave the sugar and yield the active natural products[111, 173]. Given the lack of the aglycone, satosporin C, in fermentations of *K. griseola*, the proposal of glucosylation to achieve self-resistance may be likely. However, until the biological activity of satosporins is discovered, the involvement of the glucose moieties for bioactivity remains unknown and thus it is difficult to predict the purpose of glycosylation.

There is growing evidence that certain glycosyltransferases require the presence of an auxiliary protein usually adjacent within the biosynthetic gene cluster; these helper proteins resemble P450 enzymes, except that the active cysteine responsible for binding heme is not present. This enzyme has been characterized from the pikromycin/methymycin biosynthetic gene cluster as DesVIII which allows successful D-desosamine transfer by DesVII and analogous systems include AknT, EryCII, DnrQ and CosT in aclacinomycins, erythromycin, daunorubicin and cosmomycin, respectively[170, 174]. Given that there was a protein directly adjacent to the satosporin glycosyltransferase with C-terminal domain similar to a P450, it is possible that SatD may be an auxiliary protein to allow the glucosylation of the aglycone by SatE. Disruption of the P450 region of *satD* may identify whether this protein is involved in satosporin oxidation or glycosylation.

Although the glycosyltransferase gene *satE* was found to be clustered with the other biosynthetic genes, no NDP-sugar biosynthetic genes were found. Satosporin probably utilizes dTDP-glucose from polysaccharide biosynthesis[170]. As only one glycosyltransferase was found within the *sat* cluster, it is predicted that this enzyme incorporates both glucose moieties sequentially to give rise to the monosaccharide and disaccharide molecules satosporin B and A.

Potential regulatory and self-resistance genes within the *sat* cluster

Two genes within the *sat* cluster, *ORF8* and *ORF9* showed high protein sequence similarities to regulatory genes of other biosynthetic gene clusters and, thus, were considered to be putative cluster-situated regulators (or pathway specific regulatory proteins).

ORF8 showed 51% sequence similarity to a *Streptomyces* antibiotic regulatory protein (SARP) from *Catenulispora acidiphila.* These proteins contain a helix-turn-helix motif which binds to a heptameric DNA repeat region (TCGAGxx) to initiate transcription[175]. The only potential repeated motif found was 94 bp upstream of the start codon for *satB,* where the motifs were detected, though in the complement strand, 50 and 48 bases apart. The large distances between the motifs as well as from the -10 region are in disagreement with the usual distances of 11 nucleotide (nt) between motifs and 8 nt from the -10 region[175] and thus the DNA binding location for the protein encoded by *ORF8* remains unknown. Further evidence to support the regulatory function of ORF8 was that it contained the only in frame TTA codon of the entire cluster. This codon is very rare amongst the high GC content Actinobacteria but is usually found within secondary metabolite regulatory genes as the gene encoding this transfer RNA (tRNA), *bldA*, is only transcribed at the onset of differentiation during the stationary phase[176].

The second putative regulatory gene was *ORF9,* which showed high sequence similarity with numerous unspecific regulatory proteins, notably 80% identity to a putative regulatory gene from *Streptomyces atroolivaceus*. One specific BLAST®p hit was to TlmR3 (65% similarity), a tallysomycin regulatory protein with homology to SyrP, the syringomycin regulatory protein of *Pseudomonas syringae*[177]. A similar regulatory function is thus proposed for ORF9 and gene disruption experiments may provide evidence to support this role.

Bacterial self-resistance is often accomplished via the export of the product through intracellular efflux protein. ORF2 showed 53% sequence similarity to a protein belonging to the major facilitator superfamily transporter of *Stackebrandtia nassauensis.* These membrane-bound proteins are further divided into 18 families and carryout the efflux of various compounds

including sugars, drugs, nitrate, phosphate, nucleoside, oxalate, etc.[178] The protein encoded by *ORF2* is thus proposed to serve as an efflux protein to export satosporins to the fermentation medium, where it may be non-toxic to the producing organism. This self-resistance mechanism may be used in combination with glucosylation as mentioned previously.

Other *sat* genes

No putative role in satosporin biosynthesis could be attributed to *ORF2-ORF7* and *ORF10*, which showed sequence similarity to NRPS, PKS and genes of unknown function. Their involvement in satosporin biosynthesis should be assessed by gene disruption experiments.

4.4 CONCLUSION

The data generated from this research revealed that the *sat* cluster described here is involved in the production of satosporins in the actinomycete *K. griseola* MF730-N6. The entire set of genes necessary to propose a plausible biosynthesis of satosporin A and B are present within this cluster: two PKSs for the biosynthesis of the polyketide backbone, two putative P450 monoxygenases to carry out oxidation reactions, a glycosyltransferase for the transfer of one or two glucose molecules to the macrolide, two putative regulatory genes to control the timing of secondary metabolite production and one putative self-resistance gene. The PKSs domain organization suggests that a spontaneous intramolecular cyclization via a Diels-Alder mechanism gives rise to the unique satosporin structure composed of an octalactone ring fused to an oxo-decalin moiety.

As large amounts of bacterial genomes become available, the ability to detect (and thus "turn on") cryptic secondary metabolite biosynthetic gene cluster relies on a detailed understanding of natural product biosynthesis[177]. It is thus important to keep increasing the pool of known biosynthetic genes and their products. This research represents the first step towards a functionally annotated satosporin biosynthetic gene cluster. Experiments are currently underway to confirm the involvement and possibly gain functional information of various genes within the *sat* cluster via gene deletion.

CHAPTER 5 - DRAFT GENOME SEQUENCE OF *KITASATOSPORA GRISEOLA* MF730-N6 AND INSIGHTS INTO THE POTENTIAL FOR SECONDARY METABOLITE PRODUCTION

Part of this chapter contains material previously published as:

Arens, J.C., B. Haltli, and R.G. Kerr, Draft genome sequence of *Kitasatospora griseola* strain MF730-N6, a bafilomycin, terpentecin, and satosporin producer. *Genome Announc*, 2015. **3**(2): e00208-15

5.1 INTRODUCTION

Until very recently, a single organism within the genus *Kitasatospora*, *K. setae* NBRC 14216, had its genome sequence published[136]. Its genome was 8.7 Mb in length, shared numerous developmental genes with streptomycetes and contained 24 secondary metabolite biosynthetic gene clusters, most of which had unknown products[136]. Given the affordability of sequencing genomes with current technologies such as second (by Illumina®, Roche and Life Technologies) and third generation sequencing (by Pacific Biosciences®) and the plethora of information they provide[35, 36], genome sequencing is becoming a feasible method to determine the nucleotide sequence of secondary metabolite biosynthetic genes clusters while simultaneously obtaining genome data for that organism.

The presence of numerous secondary metabolite gene clusters within a single genome is a feature shared by actinomycetes and has prompted the sequencing of numerous streptomycetes and other natural product producing Actinobacteria. This past summer, *Kitasatospora* genomic data has grown as three new genomes were published: *Kitasatospora* spp. MBT63 and MBT66[179] and *K. cheerisanensis*[180] as well as other genomes which have not yet been annotated or published but are listed as assemblies in the NCBI database such as *K. phosalacinea* NRRL B-16230, *K. setae* NRRL-16185, *K. phosalacinea* NRRL-16228, *K. arboriphila* NRRL-24581, *Kitasatospora* sp. B-11411, *K. mediocidica* KCTC 9733 and *K. azatica* KCTC 9699.

The discovery of novel natural products via genome mining is an alternative strategy to the isolation and structure elucidation of compounds and numerous bioinformatic tools now exist to rapidly annotate genomic data and predict which belong to secondary metabolite biosynthetic gene clusters[181]. Given the large number of secondary metabolite gene clusters within kitasatosporae and the low abundance of available genomic data, there is much to be gained from exploring the genome of members of the genus *Kitasatospora*.

In order to locate and analyse the genes responsible for the production of the newly isolated polyketides, satosporins, (see Chapter 3), the entire genome of *K. griseola* MF730-N6

was sequenced and not only provided the *sat* cluster (see Chapter 4), but also enabled the analysis of the remainder of its genome. The objectives of this study were thus to analyse the genome of *K. griseola* MF730-N6 and assess whether it shared similar developmental and secondary metabolite genes as previously reported for other kitasatosporae.

5.2 MATERIALS AND METHODS

5.2.1 Strain information

Kitasatospora griseola strain MF730-N6 was obtained from the International Patent Organism Depositary (IPOD) in Nashihara, Japan, under the accession number FERM BP-1045 (transferred from FERM P-No. 8247)[88]. The strain had been isolated from soil near the Umawatashi Bridge across the Iwaiga River in Nara Prefecture, Japan, and was shown to produce the diterpene terpentecin[67, 88] and, more recently, the polyketide satosporin (see Chapter 3)[129].

5.2.2 Genome sequencing and assembly

Genomic DNA was isolated from 10 mL *K. griseola* cultures grown overnight in seed medium containing 10 g/L glucose and 10 g/L yeast extract at 30°C and 200 rpm using the Qiagen genomic tip 100/G kit. The sample was kept at 4°C and sent the following day to the McGill University and Génome Québec Innovation Centre for library preparation and sequencing using the Pacific Biosciences® RS II platform using 8 SMRT® cells. DNA quality and concentration were assessed using a NanoDrop (Thermo Scientific™) spectrophotometer and the Quant-it™ PicoGreen® dsDNA assay kit in triplicate (Life Technologies), respectively. *De novo* assembly of corrected reads was performed by McGill University and Génome Québec Innovation Centre using the Hierarchical Genome Assembly Process 2 (HGAP2) analysis pipeline[182, 183].

Gap closure was achieved by first performing a *de novo* assembly using the 15 contigs generated from HGAP2 and the Geneious assembler. Primers (Table 5.1) were then generated to

amplify 3-9 kb overlapping regions between adjacent contigs using either KOD Hot Start DNA Polymerase (VWR) or EconoTaq® plus green (Lucigen); cycling conditions were as follow: 95°C for 2 min, 25 cycles of 95°C for 20 s, X°C for 10 s, 70°C for Y time and a final extension of 70°C for 5 min (see Table 5.1 for X and Y parameters). Amplicons were then sequenced using primers located ~1 kb apart (Table 5.2) to confirm the ends were consistent with the predicted assembly sequences.

5.2.3 Genome annotation and analysis

Annotation was performed by submitting the contigs generated during the assembly to the rapid annotation using subsystem technology (RAST) server and results were visualized within the SEED environment of the National Microbial Pathogen Data Resource (NMPDR) (http://rast.nmpdr.org/)[184]. Geneious was also used to visualise genome annotations[185]. Contigs were also submitted to antiSMASH to determine the biosynthetic gene clusters present within the draft genome[38]. At the time of writing, four kitasatosporae genomes were published; the genome sequences for *Kitasatospora setae* KM-6054 (AP010968)[136], *Kitasatospora cheerisanensis* KCTC 2395 (JNBY01000000)[180], *Kitasatospora* sp. MBT63 (JAIZ01000000)[179] and *Kitasatospora* sp. MBT66 (JAIY01000000)[179] were thus also submitted to RAST and antiSMASH for direct genomic comparisons. Genome alignments were performed using the progressiveMauve algorithm within the Geneious plugin[186] and dot plots were generated with the genome comparison tool within the SEED environment.

5.2.4 Accession numbers

This genome sequencing project has been deposited to GenBank under accession number JXZB00000000.

5.3 RESULTS AND DISCUSSION

5.3.1 Sequencing and assembly

Genomic DNA was extracted using the Qiagen genomic tip since gravity, rather than centrifugation, was used to purify gDNA via column chromatography. Spectrophotometric

Table 5.1. Primers and polymerase chain reaction conditions used to generate amplicons confirming overlapping contigs generated during Geneious *de novo* assembly.

Gap	Amplicon size (bp)	Primer name	Sequence 5'→3'	X (°C)	Y (min:s)	DNA polymerase
1	7815	Gap_40to44_F Gap_40to44_R	CTACCTGGCCGACTACCTCT TCAAGGTGATGGAGATGGGC	62.1	3:45	KOD
2	8042	Gap_44to45_F Gap_44to45_R	ATCCGGCAGTACGTCGTCCA AGTTGCTGGAGAGCCTGGAC	64.4	3:45	KOD
3	9000	Gap_34to42_F Gap_34to42_ R	CGCACCAACTGGACGTCGTA AGCTGGTGGTAACCCAATCA	60.4	3:45	KOD
4	8715	Gap_42to46_F Gap_42to46_ R	GGAGGGTCGGTCGGAAGGCT AGCTTCGCCAGCGTCGGCAC	68.6	9:00	EconoTaq®
5	7403	Gap_46to36_F Gap_46to36_ R	GTCCGCCAGCCCGTCGACAT AGCACCCGGTCGATCTCCGC	68.6	9:00	EconoTaq®
6	7735	Gap_36to39_F Gap_36to39_ R	GCCTCGGCAAGCTCGATGAA TGCCGAAGTACACGACCGTG	64.4	3:45	KOD
7	3338	Gap_39to38_F Gap_39to38_ R	AACAACGCGTGGCTGACCCT ACTGCCGGAGGCTCGGTGAC	64.5	1:30	KOD

Table 5.2. Sequencing primers used during gap closure experiments to confirm the identity of amplicons obtained from overlapping regions of adjacent contigs in the assembly.

Primer	Sequence 5'→3'	Note[a]
1_1_754378F	CTACCTGGCCGACTACCTCT	U
1_2_755607R	GGTCCGCATCCAGGTGTCGC	O
1_3_756582R	ACTGGGTTGATAGGCCGGAT	R
1_4_759500F	CCCCCGGATCAAAGCTCGGT	R
1_5_760606F	AACACCTAGTTCCCAACGTT	R
1_6_761413F	TGAATGCTTCCACGAAAGAG	R
1_7_762194R	TCAAGGTGATGGAGATGGGC	U
2_1_1346919F	ATCCGGCAGTACGTCGTCCA	U
2_2_1350218R	CGGAGATGGTCGGTTGGCTG	U
2_3_1349440R	GGCTTCGCCAACCCGCAGTC	U
2_4_1350731F	TGTGGCCGTAGTTGAGGATC	U
2_5_1352060F	CGCCGCCGAACTTCTCCGCC	U
2_6_1352930F	TCCCACTTCGGCCACTCGGT	U
2_7_1354966R	AGTTGCTGGAGAGCCTGGAC	U
3_1_3115313F	CGCACCAACTGGACGTCGTA	O
3_2_3116207R	GGCTTCCCTGTGGCCTTGTC	O
3_3_3123386F	GAGTCCGAGGTTCAGCAGTG	U
3_4_3124029F	ACCGCCTGGACTCCCGACAG	U
3_5_3124311R	AGCTGGTGGTAACCCAATCA	U
5_1_5517645F	GTCCGCCAGCCCGTCGACAT	U
5_2_5518516F	CGGGTCGAGGTGGCCGGGGA	R
5_3_5519328F	ATATGTTGACATCCCCGGCC	R
5_4_5523104F	CGGGTGAACTGTGGTTAAGG	R
5_5_5523911F	GGATAACAGGCTGATCTTCC	R
5_6_5524498F	ACCCGGAAGCTAAGCCTCAC	R
5_7_5525048R	AGCACCCGGTCGATCTCCGC	U
7_1_7918230F	AACAACGCGTGGCTGACCCT	U
7_2_7919688F	ATGGAGTGCGGGGAGACGGT	U
7_3_7919966R	CCAGTCCATCCCCTGTTTGT	U
7_4_7920742F	CCCGTCGATCTCGTGAACGT	U
7_5_7921568R	ACTGCCGGAGGCTCGGTGAC	U

[a]U:unique sequence, R:located within rRNA operon, O:present at other location(s) within the genome

measurements of the resulting gDNA were 1.89 and 1.80 for 260/280 and 260/230 ratio respectively, suggesting the DNA was of good quality and contained no protein contaminants. DNA concentration using PicoGreen® was calculated to be 80.36 ±1.91 ng/µL (average ±1 standard deviation).

500 µL of the fresh *K. griseola* gDNA was sent to the McGill University and Génome Québec Innovation Centre to be sequenced using Pacific Biosciences® technology. From the 150,000 zero-mode waveguides within each of the eight SMRT® cells, empty ZMWs ranged between 6.81 and 24.66% while those loaded with two enzyme complexes or other detrimental conditions ranged between 16.63 and 33.72%. ZMWs loaded with a single enzyme complex, giving rise to usable sequence reads, ranged between 57.70 and 61.91%. Average raw read quality was between 0.82 and 0.83.

Each sequencing SMRT® cell generated average raw read lengths of 4,331 bp which included sequence data from both insert and adapters. The sequencing facility processed and filtered these reads to remove the SMRTbell™ sequencing adapters, poor quality regions, and reads of insufficient length. Good quality, adapter-free sequence data were saved as continuous long reads (CLRs), also referred to as subreads. When the insert DNA between two SMRTbell™ adapters was sequenced at least twice in the zero-mode waveguide, the raw read was used to generate a high-accuracy circular consensus sequences (CCS) read.

The HGAP2 analysis pipeline was then used to generate a highly accurate (>99.99%) polished assembly in three stages[182]. Since PacBio sequencing data are only ~86% accurate, the purpose of the first stage was to correct most of the sequencing errors. Subreads were first grouped into two categories based on read length: the 219,016 subreads longer than 3 kb were used as "seed subreads" to which all the remaining subreads were aligned. The resulting consensus sequences were saved as long corrected subreads. The second stage involved using only the corrected subreads generated in the first stage to construct an assembly using Celera Assembler. In the third stage, all of the initial subreads were aligned to the assembly generated

in stage 2 using Quiver, correcting the few errors that were still present in the first assembly.

Sequencing of eight SMRT® cells generated 1,377,430 subreads (2,920,331,552 bases) and 77,677 CCS (175,515,662 bases), totalling 3,095,847,214 bases and 324- and 18-fold coverage, respectively (assuming a 9 Mb genome). 219,016 subreads were greater than 3 kb, totalling 1,065,806,608 bases and a 118X coverage assuming a 9 Mb genome. A summary of the raw data is provided in Table 5.3. The polished assembly from *K. griseola* MF730-N6 sequencing data contained 15 contigs (Quiver_34-48) ranging in size between 2.8 kb and 2.3 Mb. BLAST®n results from the 16S rRNA gene from the genome confirmed that the gDNA originated from *K. griseola* as the sequence shared >99% nucleotide identity with 16S rRNA genes from various *K. griseola* strains such as NBRC 14371 and JCM 3339.

Gap closure

To reduce the number of contigs forming the *K. griseola* MF730-N6 draft genome, Geneious was used to assemble the 15 contigs provided by the sequencing facility. The assembly was able to assemble 13 of the contigs (Quiver_37 and _43 were not part of the assembly) (Figure 5.1).

Primers were designed to amplify 7 overlapping regions, but only amplicons for gaps 1, 2, 3, 5 and 7 were obtained; the amplicon sizes were as expected. Sequences from gaps 1, 2, 5 and 7 assembled with the reference sequences of the assembly and not only confirmed the close proximity of adjacent contigs, but also allowed the correction of errors located at the extremities of the contigs. Discrepancies between overlapping contigs were probably a result of poor error correction during HGAP2 assembly due to insufficient alignment overlap. The inability of HGAP2 to assemble adjacent contigs can also be explained for gaps 1 and 5 due to the presence of rRNA operons at these extremities.

Although Geneious had predicted that the 3’ end of Quiver_34 was adjacent to the 3’end of Quiver_42 (Figure 5.1), sequencing of gap 3 amplicon revealed a complicated scenario. Quiver_43, which was not assembled by Geneious, had an identical 2 kb region to the 3’

Table 5.3. Read data generated from each of the eight SMRT® cells used to generate the data assembled into 15 contigs.

	Raw reads		Subreads		CCS		Longest >3 kb subreads	
Run	**average read quality**	**average read length (bp)**	**count**	**average read length (bp)**	**count**	**average read length (bp)**	**count**	**average read length (bp)**
1	0.83	4,422	156,513	2,376	11,001	2,494	31,655	5,174
2	0.82	3,883	169,239	1,982	6,433	2,138	25,415	4,612
3	0.82	3,923	174,754	1,995	6,911	2,151	26,340	4,620
4	0.83	5,066	190,130	2,216	13,054	2,279	30,717	5,063
5	0.83	4,928	189,057	2,207	12,417	2,279	30,607	4,983
6	0.82	4,136	162,358	2,065	8,917	2,189	24,704	4,797
7	0.83	4,102	167,765	2,057	9,269	2,215	25,008	4,772
8	0.83	4,191	167,614	2,062	9,675	2,209	24,570	4,772
Average	0.83	4,331		2,120		2,244	4,849	4,849
Sum			1,377,430		77,677		219,016	
Total bases			2,920,331,552		175,515,662		1,065,806,608	

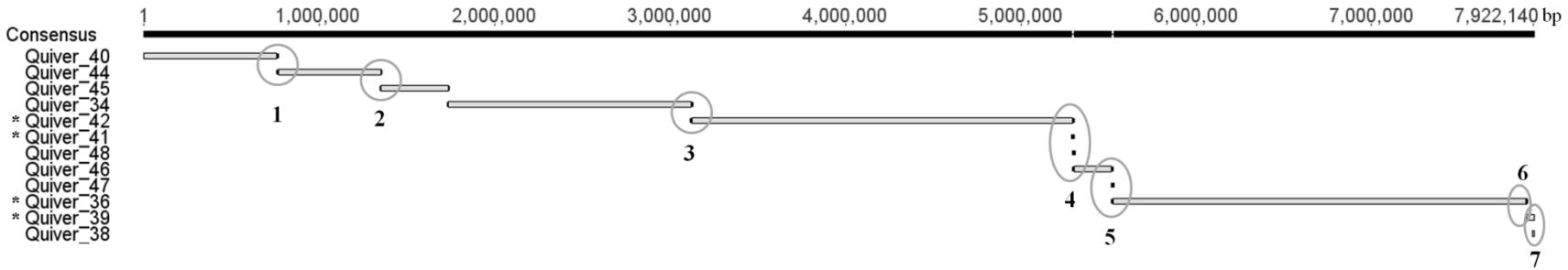


Figure 5.1. Geneious assembly using 13 of the 15 contigs provided by the sequencing facility, contigs Quiver_37 and _44 we not used in the assembly and gaps for which primers could be designed are circled and numbered according to the amplicon number, *denotes contigs assembled using the complimentary strand.

extremities of both Quiver_34 and _42, encoding various tRNA genes and a phage integrase. Quiver_34 and _42 were however concluded to be adjacent without including Quiver_43 for two reasons. First, although Quiver_43 had a 2 kb region with high identity to the 3' end of Quiver_34 and _42, the remaining 3 kb of this contig did not align with Quiver_34, suggesting that either quiver_43 was misassembled or represented another region within the genome containing the same tRNA genes along with the phage integrase (Figure 5.2). Secondly, node 167, obtained from Illumina® sequencing mentioned in Chapter 4 (Section 4.2.2), was large enough to cover the entire overlapping region between Quiver_34 and _42 while extending a further 3.3 kb onto Quiver_42 and 2.7 kb onto Quiver_34 (Figure 5.3), suggesting that these contigs were adjacent in the genome. Similarly, nodes 468 and 1430 from Illumina® sequencing confirmed junctions between Quiver_44 to _45 and Quiver_46 to _36, respectively. Gap closure via PCR thus permitted the closure of 5 gaps, resulting in 8 contigs rather than 15 and ranged in size from 5.9 kb to 3.5 Mb with a contig N50 of 2,590,787.

5.3.2 General features of the *K. griseola* MF730-N6 genome

The *K. griseola* draft genome was composed of eight contigs covering 7,966,157 bp and had an overall G + C content of 72.7%, which is the shortest genome and lowest G + C content of the four other *Kitasatospora* genomes published (Table 5.4). Genome annotation by RAST revealed a total of 7225 protein coding sequences (PEGs), 100 tRNA genes and 40 rRNA genes forming 9 complete and 5 incomplete rRNA operons[187]. As illustrated in Figure 5.4, the tRNA genes were located mostly within contig_4, while rRNA operons were distributed evenly throughout the contigs.

Of the 7225 coding DNA sequences, 2540 were genes encoding hypothetical proteins while 4685 were assigned putative functions. 2170 (31%) were categorized into 425 subsystems and, when compared to those determined by RAST for *K. setae*, *K. cheerisanensis*, *Kitasatospora* spp. MBT63 and MBT66 (Figure 5.5), showed a comparable distribution of

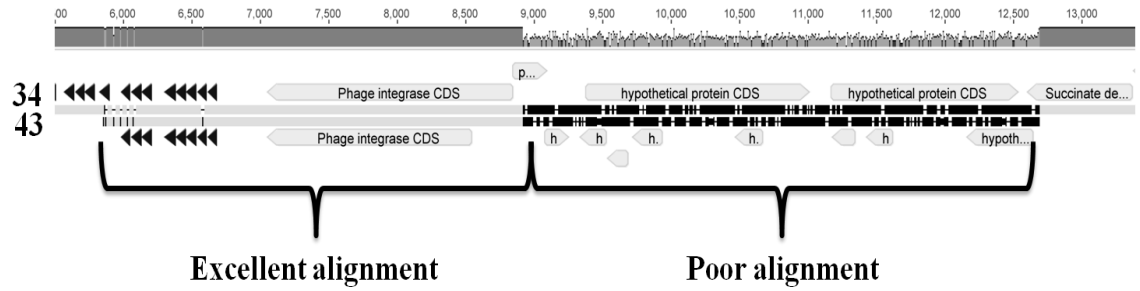


Figure 5.2. Alignment of the 3' end of Quiver_34 with Quiver _43 illustrating how even though there was a 2 kb region with excellent nucleotide alignment, the remainder of quiver_43 failed to align to Quiver_34.

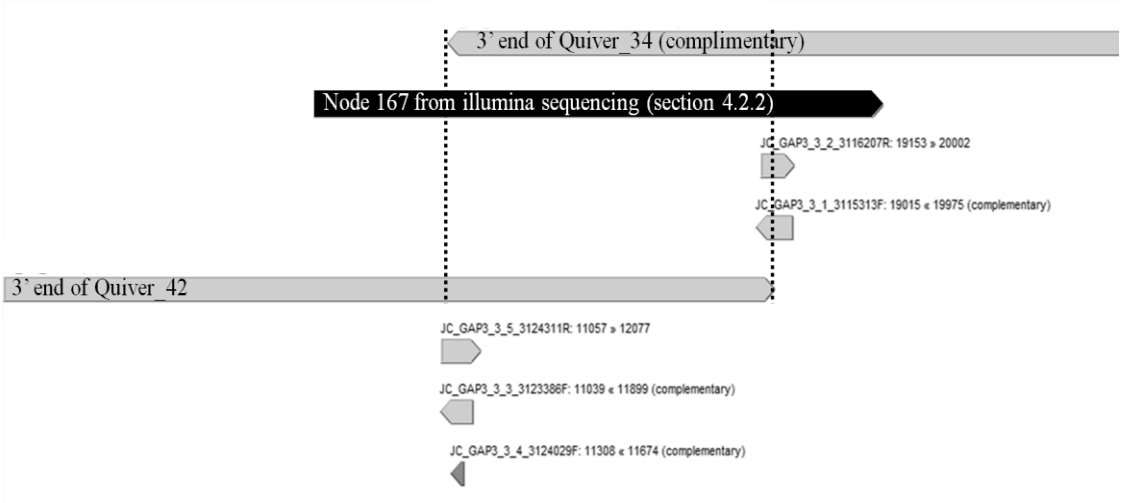


Figure 5.3. Evidence suggesting that contigs Quiver_42 and Quiver_34 are adjacent on the genome of *K. griseola* includes node 167 (black bar) obtained from Illumina® sequencing (see Section 4.2.2) which covered the entire overlapping region between the two contigs (grey) with high nucleotide identity; contig overlapping extremities are shown by dotted vertical lines and alignment with gap 3 amplicons is also shown.

Table 5.4. General features of *K. griseola* MF730-N6 genome compared to the other published kitasatosporae genomes.

Strain	*K. griseola* MF730-N6	*K. cheerisanensis* KCTC 2395	*Kitasatospora* sp. MBT63	*Kitasatospora* sp. MBT66	*K. setae* KM-6054
Total length (bp)	7,966,157	8,035,179	9,900,000	10,400,00	8,783,278
Contigs (no.)	8	5 scaffolds (178 contigs)	849	45	1
G + C content (%)	72.7	73.6	73.0	73.2	74.2
# RNA genes	140[a]	NR	78	81	101[b]
# tRNA genes	100	72	NR	NR	74
# rRNA operons	9 complete + 5 incomplete	9	NR	NR	9
# Protein encoding sequences (PEG)	7,226	7,810	8,651	8,827	7,569
Accession number	JXZB00000000	JNBY00000000	JAIZ00000000	JAIY00000000	AP010968
Published		June 2014	June 2014	June 2014	November 2010
Reference	This study	180	179	179	136

[a] probably an overestimation due to rRNA and tRNA genes at numerous contig boundaries, [b]provided by RAST analysis rather than from publication; NR: not reported, no.: numbers.

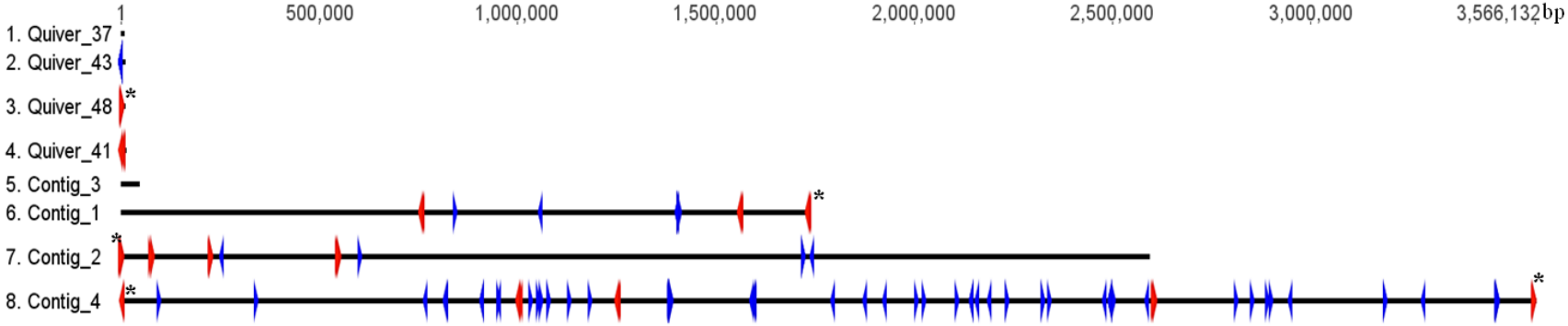


Figure 5.4. Distribution of rRNA operons (red) and tRNA genes (blue) within the genome of *K. griseola* illustrating that numerous incomplete rRNA operons (indicated by stars) were located at the extremities of contigs which probably interfered with the assembly.

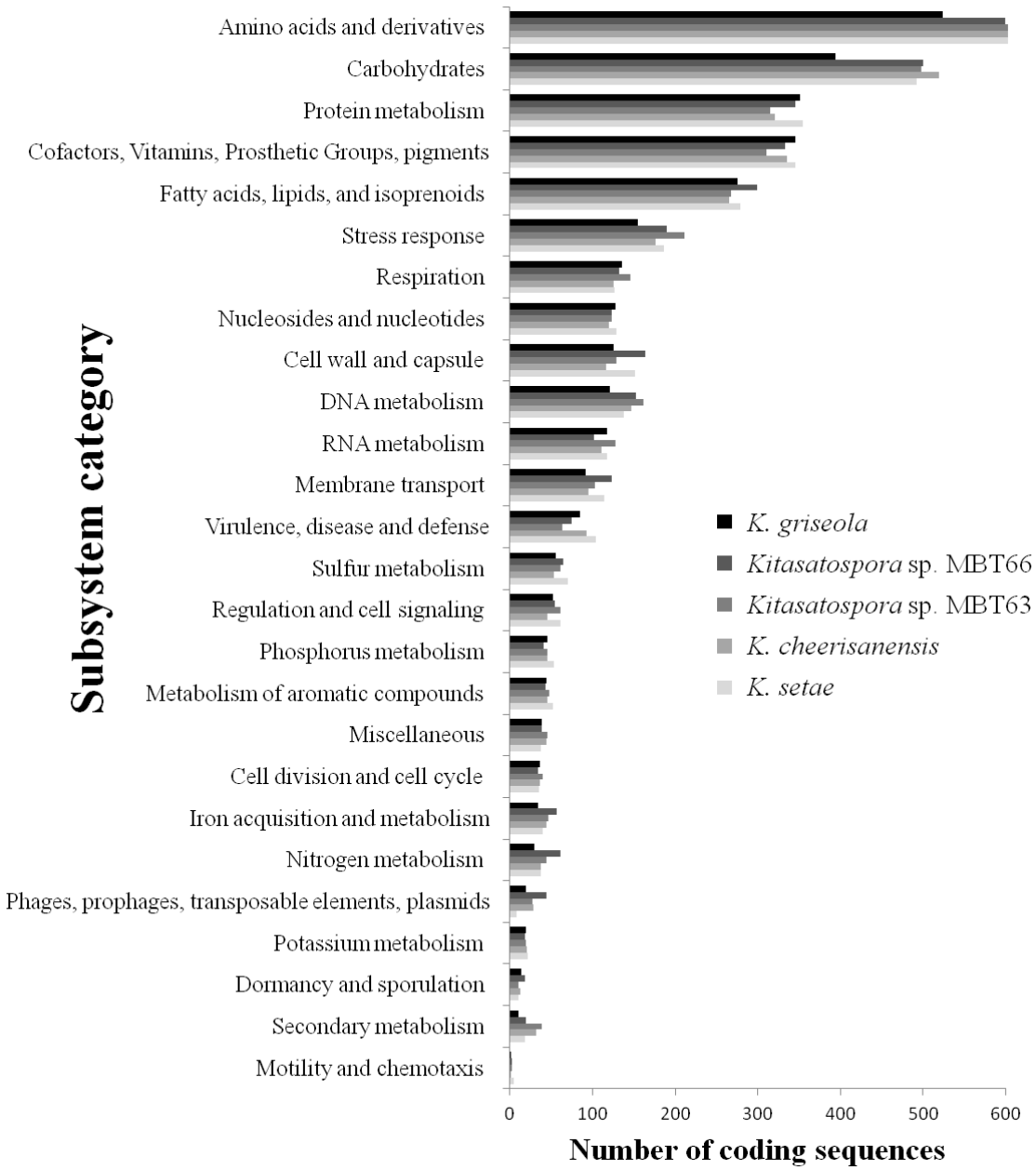


Figure 5.5. Number of coding sequences grouped into each subsystem category by RAST illustrating a comparable detection of genes between all published kitasatosporae genomes.

coding sequences (CDSs) into subsystem categories. Although other kitasatosporae have recently been sequenced according to NCBI, their genome sequence was not yet available and, thus, comparison with these organisms was not possible.

For all kitasatosporae, the majority of coding sequences belonged to the amino acid and derivative subsystem category, involving functions related to amino acid biosynthesis and catabolism; *K. griseola* had 524 CDSs in this category. Genes related to carbohydrate catabolism were also abundant in *K. griseola* (394 CDSs), including genes in pathways such as glycolysis, pentose phosphate pathway, Entner-Doudoroff pathway, the tricarboxilic acid (TCA) cycle and various anabolic and catabolic processes of sugars such as lactose, galactose, mannose and glycogen. The third most prevalent subsystem category was protein metabolism with 351 coding sequences in *K. griseola.* This category included tRNAs and genes involved in their aminoacylation as well as genes for protein chaperones, ribosome biogenesis, ribosomal proteins and protein processing, modification and degradation.

Alignment of the *K. griseola* genome with that of *K. setae* and *K. cheerisanensis* using both progressiveMauve and the SEED comparison tool to generate the dot plot in Figure 5.6 (A, C) and (B, D), respectively, revealed that there was a high level of synteny with both genomes as would be expected for closely related organisms. As gap closure for both *K. griseola* and *K. cheerisanensis* genome was not completed, the order and orientation of the 8 contigs and 5 scaffolds, respectively, could not be determined. It is thus possible that genome rearrangements, which often occur at the terminal regions of the genome, were not detected here[188]. It can however be concluded that the core portion of the genome (contig_4 in *K. griseola*) is highly syntenic to that of *K. setae*.

Since no telomere-associated proteins homologous to KSE_73020 and KSE_73030 encoding Tap and Tpg, respectively, could be detected, it was not possible to identify the terminal sequences of the genome. Furthermore, no sequences at any contig extremities contained the conserved sequence CCCGCGGAGCGGG located at the start of many

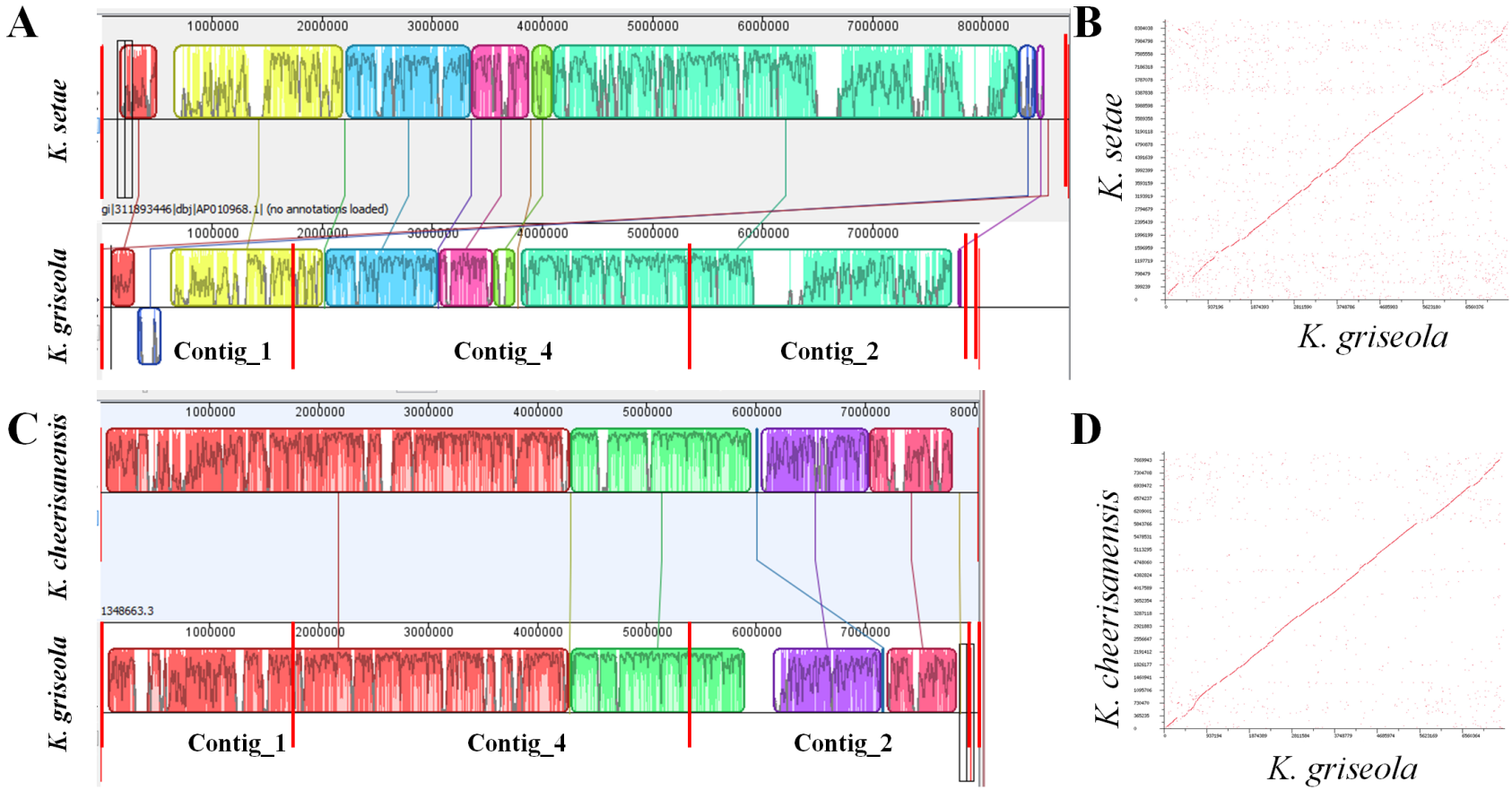


Figure 5.6. Genome synteny between *K. griseola* and *K. setae* **(A and B)** and *K. griseola* and *K. cheerisanensis* **(C and D)** using progressiveMauve and the SEED illustrating that there was high synteny between the genome of the newly sequenced organism to that of previously sequenced members of the same genus; each coloured block represents a region without rearrangements and is connected with its orthologous block, *K. griseola* contigs were arranged in the order: 1, 4, 2, 3, 37, 41, 43 and 48 and *K. cheerisanensis* contigs were kept in ascending order in order to simplify visualization of the alignments, red vertical lines represent the contig boundaries in *K. griseola*.

Streptomyces terminal sequences as well as in that of *K. setae*[189]. *Streptomyces* and *K. setae* chromosomes are linear with large inverted repeat terminal regions. *Kitasatospora setae* has at its extremity a 127 kb sequence that is inverted and repeated at the other end[136]; these large repeated sequences are difficult to resolve during *de novo* assembly. Until the remainder of the gaps are closed, the features of the terminal sequences of *K. griseola* remain unknown. Interestingly, Tap and Tpg homologs were identified in *Kitasatospora* spp. MBT63 (PEG.8582, 64% aa identity and PEG.8584, 57% aa identity) and MBT66 (PEG. 2870, 72% aa identity and PEG.2869, 69% aa identity), respectively but both were absent in the genome of *K. cheerisanensis*.

The putative bidirectional origin of replication *oriC* was located at bases 2,313,299-2,314,082 (784 bp) of contig_4 and shared 86% nucleotide identity with *oriC* of *K. setae* located at bases 4,370,854-4,372,117 (1,264 bp). G + C content of the putative *oriC* region was 62.2%. It was flanked by *dnaA* and *dnaN* which encode for the chromosomal replication initiator protein DnaA and DNA polymerase III beta subunit DnaN, respectively. *OriC*, spanning the *dnaA-dnaN* intergenic region, also featured numerous nonameric DnaA boxes to which the DnaA protein would bind and unwind the DNA to allow replication; some streptomycetes have been reported to have 19 DnaA box-like sequences, while 20 are reported for *K. setae*[136, 190]. Sixteen DnaA box sequences using the sequence (T/C/G)(T/C)(G/A/C)TCCAC(A/C/G) were located within the *oriC* region in *K. griseola.* Alignment with putative *oriC* regions from four other kitasatosporae revealed that most of the DnaA boxes were also present at the same locations (Figure 5.7). Furthermore, 36 of the 40 "strong" genomic DnaA box sequences (TT(A/G)TCCACA) were located within contig_4 (Figure 5.8), suggesting that, along with the presence of *oriC*, Contig_4 contained the core portion of the genome[191]. Although *Streptomyces* and *Kitasatospora* are very similar genera, a phenotypic characteristic that distinguishes them is the presence of the *meso* isomer of diaminopimelic acid (DAP) in the substrate mycelial peptidoglycan layer[86]. Genetically, *dapF* encodes for the diaminopimelate epimerase required to convert the LL-DAP

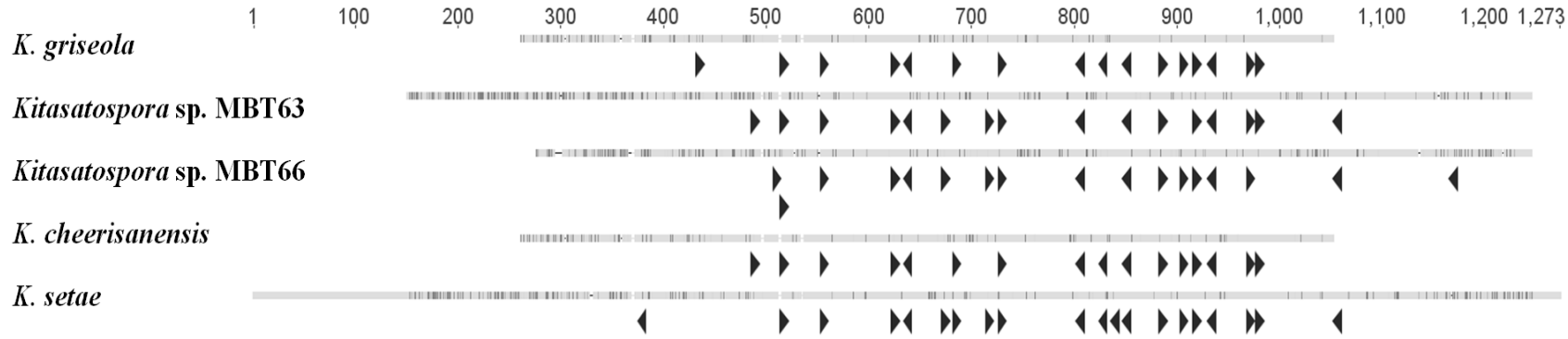


Figure 5.7. Alignment of the putative *oriC* regions from five kitasatosporae genomes illustrating the presence of DnaA box-like sequences ((T/C/G)(T/C)(G/A/C)TCCAC(A/C/G)) shown as black arrows at similar locations within *oriC.*

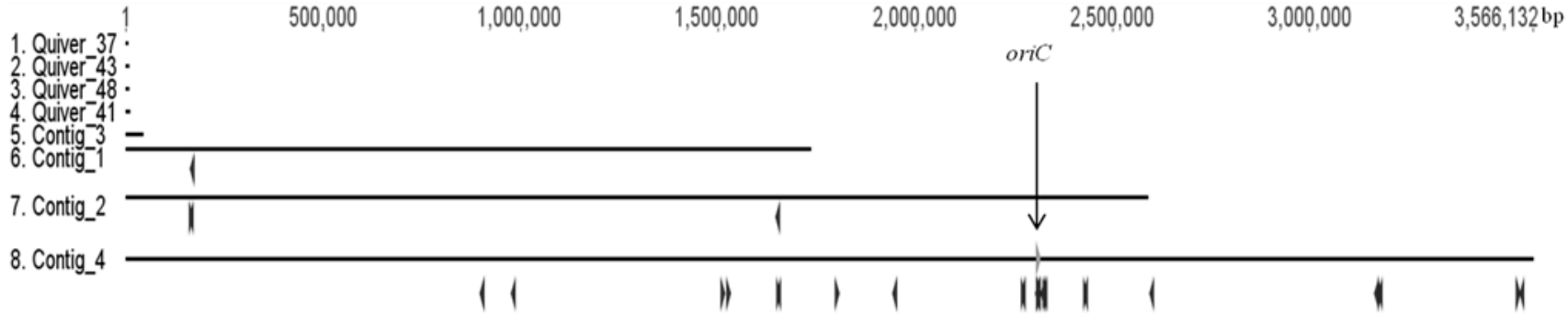


Figure 5.8. Distribution of "strong" DnaA box sequences (TT(A/G)TCCACA) (black arrows) within the genome where most of them were located on contig_4, surrounding the labelled origin of replication *oriC* (light grey arrow)

into *meso*-DAP. Although three *dapF* paralogs were identified from *K. setae*; this may be the exception rather than the rule for this genus as only one homolog was identified from *Kitasatospora* spp. MBT63 and MBT66[179] as well as from the genomes of *K. griseola* and *K. cheerisanensis* where KGR_1687 shared 88% aa identity and KCH_5206 shared 89% aa identity with KSE_53750 (*dapF)*, respectively. The protein responsible for adding DAP to *N*-acetylmuramyl-L-alanyl-D-glutamate during peptidoglycan biosynthesis is MurE, a UDP-*N*-acetylmuramoylalanyl-D-glutamate-2,6-diaminopimelate ligase was also located within the genome of *K. griseola* (KGR_4366) and shared 88% aa similarity to KSE_21280 and conserved gene synteny. All genes within the *dcw* cluster required for division and cell wall were present in *K. griseola* (KGR_4354-4375) as was reported for *K. setae* and *Kitasatospora* MBT63 and MBT66[136, 179].

Genetic classification of organism to either *Streptomyces* or *Kitasatospora* relies on muti-locus genotyping of 16S rRNA gene, RpoB, RecA and more recently SsgB gene products[97]. In 2013, Girard et al.[97] reported that SsgB aa sequences differed by only 1 aa within streptomycetes and by 3 aa between the two genera; DNA sequences varied by ~50 nt between genera. At that time, only the genome of *K. setae* was available for comparison. Since then, the same authors confirmed that the SsgB protein from two new kitasatosporae (*Kitasatospora* spp. MBT63 and MBT66) followed the same trend[179]. The data generated here was consistent with previous conclusions as the SsgB aa sequences (Figure 5.9) were extremely conserved between members of the genus *Kitasatospora* where *K. griseola*, *K. cheerisanensis* and *Kitasatospora* spp. MBT63 and MBT66 shared almost the same SsgB protein sequence, differing in only the last residue to that of *K. setae*. *SsgB* nucleotide sequences varied by as much as 29 nt between kitasatosporae but varied by 56 to 61 nt to that of *S. coelicolor*. Using two additional kitasatosporae sequences, it appears that the paradigm holds true and that SsgB remains an appropriate taxonomic marker to distinguish streptomycetes from kitasatosporae.

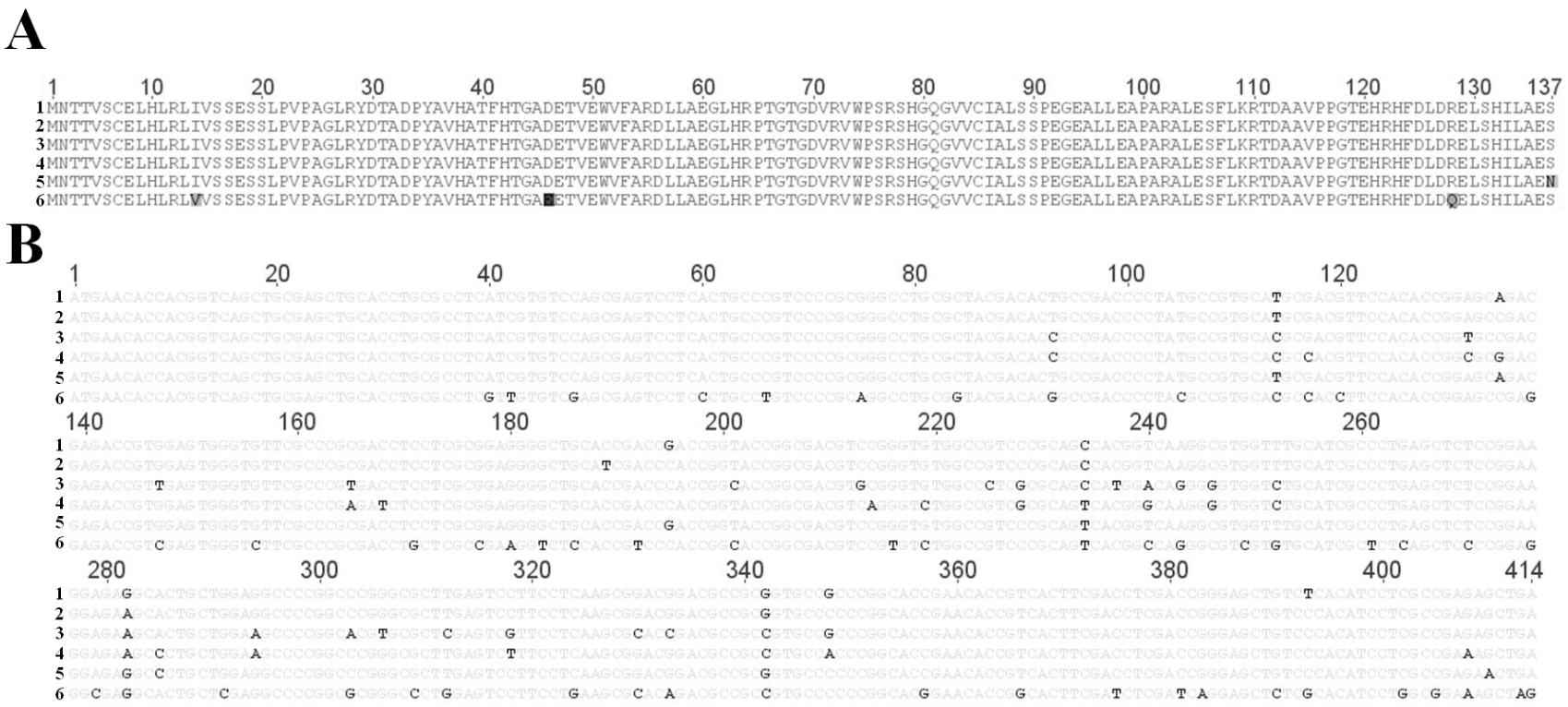


Figure 5.9. Amino acid **(A)** and nucleotide **(B)** alignment showing how highly conserved the amino acid sequence of the SsgB protein sequence is between members of the genus *Kitasatospora* compared to *Streptomyces coelicolor;* 1: *K. griseola*, 2: *K. cheerisanensis*, 3: *Kitasatospora* sp. MBT63, 4: *Kitasatospora* sp. MBT66, 5: *K. setae* and 6: *S. coelicolor;* discrepancies are highlighted in the protein sequences and bolded in the DNA sequences.

5.3.3 Genes involved in developmental regulation

Streptomycetes have an elaborate life cycle, beginning with germination of a unigenomic spore. Chromosomal replication, cell wall growth and branching result in a dense substrate mycelium which, in response to nutrient limitation and cell density, develops into aerial hyphae, each containing multiple copies of the chromosome[71]. The onset of aerial hyphae formation is controlled by numerous *bld* genes and also involves a protein encoded by *ram* genes (SapB) as well as amphipathic surface-active proteins (chaplins and rodlins)[192]. Growth and septation of aerial mycelia, controlled by *whi* genes and SsgA-like proteins (SALPs) result in long chains of spores, each containing a single copy of the chromosome[71].

Genes involved in aerial hyphae formation

Orthologs of almost all aerial mycelia regulatory genes (*bld*) were identified in *K. griseola* as was previously reported for *K. setae* and *Kitasatospora* spp. MBT63 and MBT66[136, 179] and are listed in Table 5.5. In agreement with other kitasatosporae sequenced to date, no true ortholog of *bldB* could be identified by BLAST®p in *K. griseola.* Gene synteny revealed that the genomic region (KSE_53190-KSE_53240) of all five kitasatosporae was identical except that no *Kitasatospora* genes homologous to the *S. coelicolor bldB* (SCO_5723) were present in this genomic region. *BldB* has been reported to be essential for production of aerial mycelia and secondary metabolites in *S. coelicolor* in glucose–containing media[193]; the same appears to be untrue for *Kitasatospora* as they are still capable of producing both aerial hyphae and secondary metabolites. In *S. coelicolor*, SCO_7246 has been described as a paralog of *bldB*[136] and although orthologs were located in all kitasatosporae, it is not known what function this gene has and whether it is involved in similar processes as *bldB*.

SapB, encoded by the ram cluster in *S. lividans*, and the other lanthionine-containing peptides within class 3 lantibiotics have been shown (SapT, AmfS) and suggested (labyrinthopeptins A1 and A2) to be involved in aerial mycelia formation in streptomycetes[194]. In contrast to *K. setae* where no such genes could be located[136], orthologs of the *ram* gene cluster

Table 5.5. Aerial mycelia orthologs identified in *K. griseola,* number in parentheses represent percent amino acid identity to *K. setae* homolog.

	***K. griseola* MF730-N6**	***K. setae* KM-6054**	***S. coelicolor* A3(2)**
Aerial mycelium			
bldA	KGR_rna_116 (98)	KSE_t0069	SCO_t24
bldB	(KGR_1513)[a] (87)	(KSE_16220)[a]	SCO_5723 (SCO_7246)
bldC	KGR_6221 (100)	KSE_41620	SCO_4091
bldD	KGR_1201 (98)	KSE_13950	SCO_1489
bldG (=*rsbV*)	KGR_5723 (98)	KSE_35840	SCO_3549
bldH (=*adpA*)	KGR_4894 (78)	KSE_26930	SCO_2792
bldKA-bldKE	KGR_6785-KGR_6789 (86-89)	KSE_48250-KSE_48290	SCO_5112- SCO_5116
bldM (=*whiK*)	KGR_5305 (100)	KSE_31060	SCO_4768
bldN (=*adsA*)	KGR_5487 (88)	KSE_33690	SCO_3323
amfC	KGR_6370 (91)	KSE_43740	SCO_4184
Lantibiotic-like surfactant (SapB)			
ramC	KGR_3869 (46)	KSE_63000	SCO_6685
ramS	KGR_3868 (36)[b]	ND	SCO_6682
ramA	KGR_3867 (42)[b]	ND	SCO_6683
ramB	KGR_3865 (41)[b]	ND	SCO_6684
ramR	ND	ND	SCO_6681

ND: not detected, [a]represents the kitasatosporae orthologs of the *bldB S. coelicolor* paralog SCO_7246, [b]aa identity to *S. coelicolor* homolog

(*ramC, S, A and B*) were located in both *K. griseola* and *K. cheerisanensis* but not in *Kitasatospora* spp. MBT63 and MBT66. As shown in Figure 5.10 (A), the clusters contained the usual lanthionine synthetase gene (KGR_3869), the precursor peptide gene (KGR_3868) and the ATP-binding cassette (ABC) transporter system genes (KGR_3867, KGR_3865). The presence of the synthetase upstream of the precursor peptide was a characteristic feature of type 3 lantipeptides. The predicted leader and core peptide sequences are shown in Figure 5.10 (B) where dehydrated serine and threonine are shown as dehydroalanine (Dha) and dehydrobutyrine (Dhb), respectively, with two cysteine residues available for cyclization. No *ramR* ortholog was located in either kitasatosporae and, although present in both *S. coelicolor* and *S. griseus*, it was not present within the LabA1/A2 clusters of *Actinomaduras namibiensis*[194]. Since the peptide predicted to result from the orthologous genes in *K. griseola* was different from previously reported class 3 lantipeptide, the resulting peptide would require a different name.

Nine chaplin genes (KGR_5470-5473, KGR_3094-3095, KGR_3635, KGR_1661 and KGR_4395) and two rodling genes (KGR_5476-5477) were located within the genome of *K. griseola*. These amphipathic proteins are believed to assist in breaking the surface tension of the water at the surface of the colony to allow the hyphae to grow vertically[192].

Differentiation into spores (*whi* and *SALPs*)

Following the formation of multigenomic aerial hyphae, unigenomic spores are formed following glycogen deposition, septation, prespore rounding, cell wall thickening and spore pigment deposition[71, 192, 195] (Figure 5.11). The differentiation of aerial hyphae into spore chains is regulated by numerous w*hi* genes and involves many SALPs. As was reported for *K. setae,* orthologs of all *whi* genes, except *whiJ*, were located in *K. griseola* suggesting that kitasatosporae seem to regulate sporulation in a similar manner (Table 5.6). *WhiJ* (SCO4543) has recently been described as a sporulation repressor that is inactivated by a signal provided by adjacent genes encoding for an acidic protein related to *bldB* (SCO4542)[196]. This may provide an explanation why *whiJ* and *bldB* seem to be the only two genes within the *whi* and *bld* clusters to

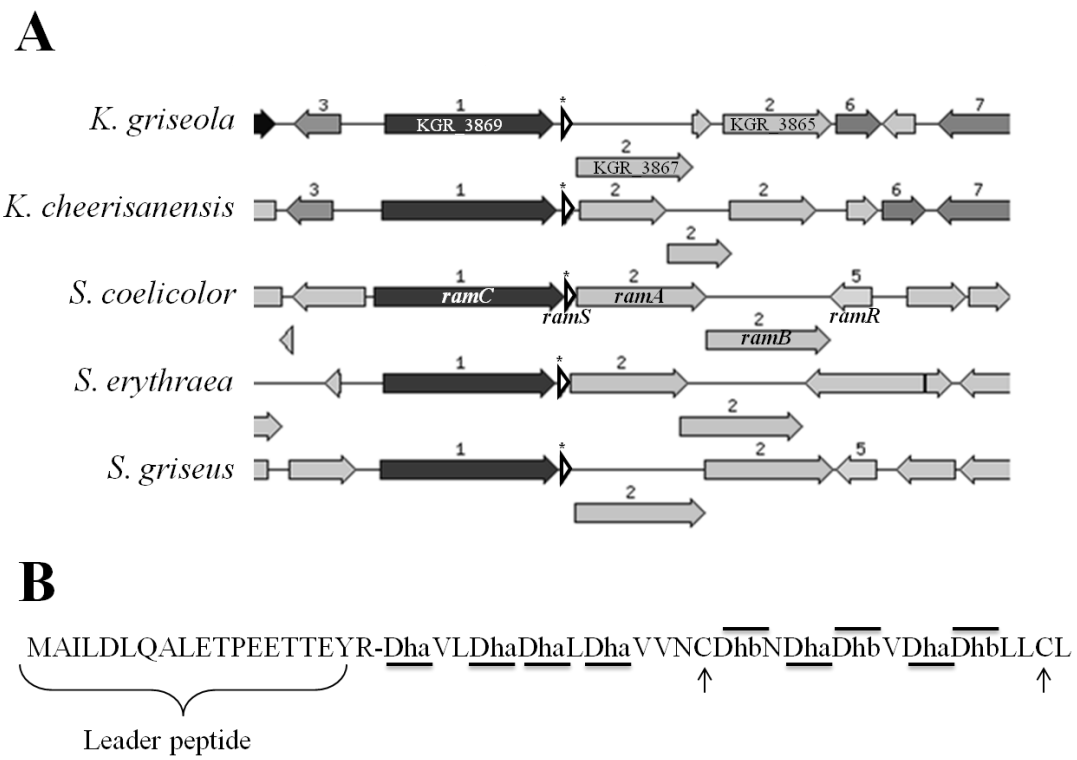


Figure 5.10. Gene synteny of the *ram* cluster between various streptomycetes and kitasatosporae, where orthologs are identified using the same numbers, highlighting similar gene organization: lanthionine synthethase (*ramC*, black, 1), precursor peptide (*ramS*, white, star), ABC transporter (*ramA/B*, grey, 2) and regulatory (*ramR*, grey, 5) genes. *K. griseola* orthologous genes are identified and *ramR* ortholog was absent **(A)** and predicted leader peptide and lantipeptide sequence following dehydration of serine and threonine residues by KGR_3869 to result in Dha and Dhb, respectively with possible cyclization between Dha/Dhb and two cysteine residues (shown with arrows) **(B)**.

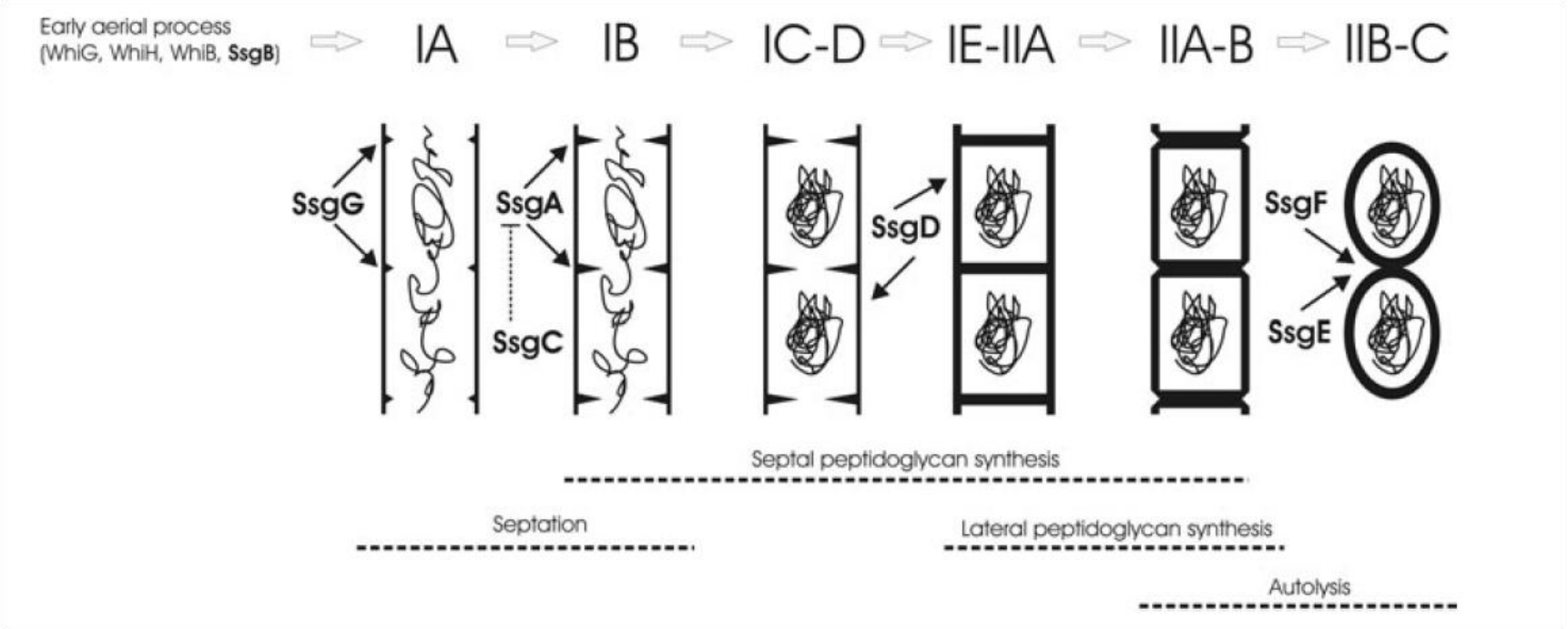


Figure 5.11. Sporulation in streptomycetes involves various *whi* and SsgA-like protein (SALP) genes, where *ssgG* is involved in controlling the site of septation, *ssgA* initiates septum formation, *ssgD* is involved in peptidoglycan synthesis while *ssgF* and *ssgE* are involved in autolysis of the mature spore chain. Reproduced, with permission, from Noens et al.[195]

Table 5.6. *Whi* and SsgA-like protein (SALP) orthologs located in the genome of *Kitasatospora griseola* along with the corresponding genes in *K. setae* and *S. coelicolor*. Number in parentheses represents percent amino acid identity to the orthologous gene in *K. setae*.

	***K. griseola* MF730-N6**	***K. setae* KM-6054**	***S. coelicolor* A3(2)**
Sporulation			
whiA	KGR_1863 (99)	KSE_55390	SCO1950
whiB	KGR_5109 (100)	KSE_29410	SCO3034
whiD (=*wblB*)	KGR_5307 (95)	KSE_31070	SCO4767
whiE	KGR_3688-KGR_3695 (82-94)	KSE_72410-KSE_72480	SCO5314-SCO5321
whiG	KGR_7183 (93)	KSE_52540	SCO5621
whiH	KGR_1723 (95)	KSE_54050	SCO5819
whiI	KGR_1834 (95)	KSE_55090	SCO6029
whiJ	ND	ND	SCO4543
sigF	KGR_5912 (88)	KSE_37310	SCO4035
sigN	KGR_5913 (89)	KSE_37320	SCO4034
crgA (=*whiP*)	KGR_6090 (90)	KSE_39340	SCO3854
SsgA-like proteins (SALPs)			
ssgR	KGR_6125 (80)	KSE_39760	SCO3925
ssgA	ND	KSE_39770/KSE_12250	SCO3926
ssgB	KGR_1264 (99)	KSE_14600	SCO1541
ssgC	ND	KSE_1860	SCO7289
ssgD	KGR_5035 (71)	KSE_28620	SCO6722
ssgE	ND	ND	SCO3158
ssgF	ND	KSE_5120	SCO7175
ssgG	KGR_5020 (97)	KSE_28490/KSE_11930	SCO2924
Other SALPs	ND	KSE_6570	
	KGR_1906 (76)	KSE_55820	
	KGR_2498 (69)	KSE_60010	
	ND	KSE_68960	
	KGR_561	ND	
	KGR_873	ND	
	KGR_2403	ND	

ND: not detected

be absent from kitasatosporae.

Besides the *ssgB* gene mentioned earlier, eight other putative SsgA-like protein (SALP) genes were found within the genome of *K. griseola* (Table 5.6) and include putative orthologs of *ssgD* and *ssgG*. Although SsgA has been shown to be required for sporulation in streptomycetes[197], no *ssgA* ortholog was found in *K. griseola*. Comparison of genomic regions at this locus(Figure 5.12) showed that although *ssgR,* which encoded an IclR family regulator, was present in *K. griseola, ssgA* was not; *Kitasatospora* sp. MBT66 did not have *ssgA* nor *ssgR* orthologs, while *Kitasatospora* sp. MBT63 had a *ssgA* ortholog but no *ssgR* ortholog. *K. griseola* had three SsgA-like protein genes to which no *K. setae* ortholog could be assigned; this may suggest that one of these genes may function as SsgA or that septation is initiated differently in members of this genus. Knock-out experiments of each SALP would be necessary to assign their function in *K. griseola* as has been done in *S. coelicolor*[195].

5.3.4 Genes involved in secondary metabolism

Analysis of the *K. griseola* MF730-N6 genome by antiSMASH (http://antismash.secondarymetabolites.org/) revealed 24 putative secondary metabolite biosynthetic gene clusters. Cluster 2 was discarded as it contained a single predicted biosynthetic gene, which following further analysis by BLAST®p and synteny search with other kitasatosporae, indicated that the genomic region was well conserved and did not encode for secondary metabolite biosynthesis; the predicted biosynthetic gene was similar to an endonuclease (91% aa identity with endonuclease from *Streptomyces* sp. NRRL S-384). A total of 23 biosynthetic genes clusters were thus present in the genome of *K. griseola* MF730-N6. Similarly, 24 and 38 secondary metabolite biosynthetic gene clusters are predicted for *K. setae*[136] and *Kitasatospora* sp. BMT66[179], respectively.

As listed in Table 5.7 and Table 5.8, the analysis predicted seven lantipeptide, six polyketide (PK), three terpene, two siderophore, two nonribosomal peptide-polyketide hybrid

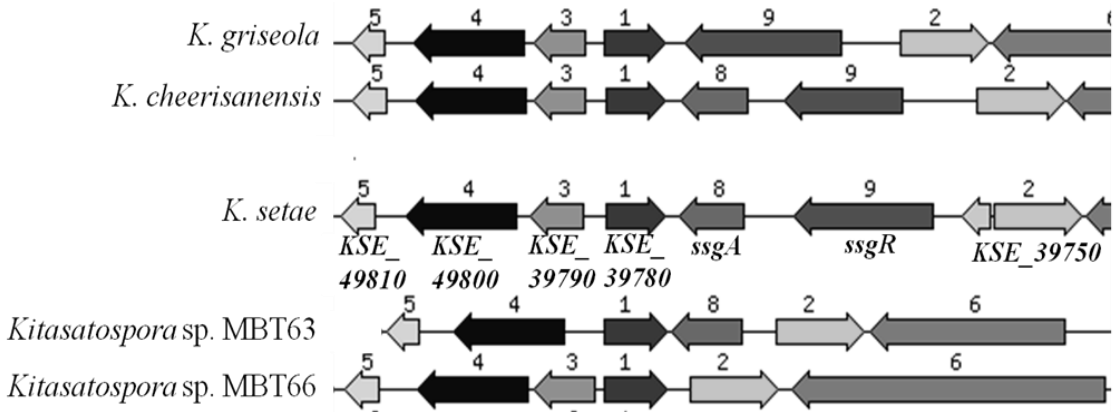


Figure 5.12. Evidence of gene synteny between kitasatosporae for *ssgR* (number 9) and *ssgA* (number 8) where a *ssgR* ortholog is present in *K. griseola*, while the *ssgA* ortholog is absent. Genes in *K. setae* are labelled for reference and orthologs are represented by the same number.

Table 5.7. Secondary metabolite biosynthetic gene clusters predicted by antiSMASH that are present and absent in the genome of *K. setae* for which the biosynthetic products are known.

Cluster[a]	Contig[b]	Location[c]	Class	ORF ID in *K. setae* (KSE_XXXXX)	Metabolite and/or propeptide prediction
Present in *K. setae*					
11	contig_4	74,391-100,828	Terpene	17490-17710	Hopanoids
19	contig_4	2,852,268-2,874,460	Terpene	46170-46000	Germacradienol/geosmin
21	contig_1	454,424-551,853	PKS (type 1)	73580-73410	Bafilomycin (setamycin)
20	contig_4	2,974,386-3,015,348	Unclassified	27320-27430	Similar to valanimycin biosynthetic genes
9	contig_2	2,333,740-2,376,204	PKS (type 2)	72290-72660	Spore pigment (encoded by *whiE* genes)
Absent in *K. setae*					
8	contig_2	1,881,136-1,902,071	Terpene	N/A	Terpentecin
24	contig_1	1,221,976-1,321,691	PKS (type 1)	N/A	Satosporin
10	contig_2	2,516,932-2,539,439	Lantipeptide[d] (class 3)	N/A	Spore-associated protein (SapB-like) (encoded by *ram* cluster)

[a]cluster number as per antiSMASH analysis

[b]contig number following gap closure as mentioned previously

[c]nucleotide of the entire cluster predicted by antiSMASH within the specified contig

[d]prepropeptide prediction: MAILDLQALETPEETTEYR-SVLSSLSVVNCTNSTVSTLLCL

Table 5.8. Secondary metabolite biosynthetic gene clusters predicted by antiSMASH that are present and absent in the genome of *K. setae* for which the biosynthetic products are not known.

Cluster[a]	Contig[b]	Location[c]	Class	ORF ID in *K. setae* (KSE_XXXXX)
Present in *K. setae*				
13	contig_4	223,839-234,639	Bacteriocin	19150-19400
14	contig_4	569,839-584,444	Siderophore	12610-12700
12	contig_4	117,317,158,363	PKS (type 3)	17920-18180
18	contig_4	2,522,507-2,544,849	Lantipeptide	42430-42240
1	contig_2	130,850-146,294	Siderophore	53760-53870
Absent in *K. setae*				
17	contig_4	1,856,553-1,908,288	PKS (type 1)	N/A
3	contig_2	720,683-777,863	PKS (type 1 & 2)	N/A
5	contig_2	871,577-920,858	NRPS/PKS (type 1)	N/A
4	contig_2	772,034-849,908	NRPS/PKS (type 1 & 2)	N/A
7	contig_2	1,673,073-1,742,609	NRPS	N/A
6	contig_2	1,221,534-1,267,236	Lantipeptide (class 2)[d]	N/A
16	contig_4	1,743,904-1,768,710	Lantipeptide (class 1)[e]	N/A
15	contig_4	681,081-703,789	Lantipeptide	N/A
22	contig_1	991,487-1,015,146	Lantipeptide	N/A
23	contig_1	1,160,634-1,183,213	Lantipeptide	N/A

[a]cluster number as per antiSMASH analysis

[b]contig number following gap closure as mentioned previously

[c]nucleotide of the entire cluster predicted by antiSMASH within the specified contig

[d]prepropeptide prediction: MSKVLELQKMVTVEGS-PIDQEAISVSSCDSHSC or MPTVVHRSDVHGA-ESATGHCRRRLRSGLGAVSERPLM

[e]prepropeptide prediction: MRNEILVQQDTDLDLDLRIAEVAEETQAFGQ-GTFTSPSSYAIGTRCPVCC

(NRP-PK), one nonribosomal peptide (NRP), one bacteriocin and one unclassified biosynthetic gene clusters. They are distributed throughout the genome (Figure 5.13). Of the 23 biosynthetic gene clusters predicted in *K. griseola*, eight had known products and ten were present in *K. setae*, indicating that although the two organisms were closely related, over half of the secondary metabolite gene clusters were unique to *K. griseola* MF730-N6. AntiSMASH results are available at http:antismash.secondarymetabolites.org/upload/689503c5-e552-414b-b32b-2e128b54e624/index.html.

Lantipeptides

Lantipeptides are lanthionine and methyl-lanthionine containing ribosomally synthesized peptides which often exhibit antibacterial activities (except class 3 lantipeptides)[194]. Seven lantipeptide biosynthetic gene clusters were predicted to be present within the genome of *K. griseola*. Only one had a known product, a peptide similar to the spore-associated protein SapB which was mentioned in Section 5.3.3, which was predicted to be involved in aerial hyphae formation and was encoded by ram orthologous genes within cluster 10 (KGR_3865-3869). The lantipeptide encoded by genes within cluster 18 (KGR_6257-6265) did not have a known metabolite but was found to have orthologs in both *K. setae* and *K. cheerisanensis* with conserved gene synteny. The five remaining lantipeptide gene clusters (clusters 6, 15, 16, 22 and 23) had unknown products and only two prepropeptide sequences could be predicted by antiSMASH, as noted in Table 5.8. Othologs in *K. cheerisanensis* were identified for all gene clusters but #6, which instead had orthology and gene synteny to a cluster in *S. griseus*. Possible *K. setae* orthologs were also detected by RAST but not by antiSMASH for lantipeptide clusters 15 and 22 (Figure 5.14). These genes were however not previously highlighted as secondary metabolite genes from *K. setae*[136].

Polyketides, nonribosomal peptides and hybrids thereof

Polyketides represent a clinically important class of natural products biosynthesized via the condensation of simple units, usually malonyl- and methylmalonyl-CoA, to starter units such

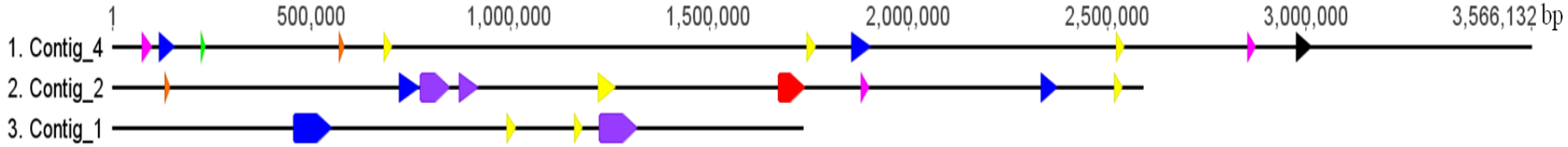


Figure 5.13. Genomic distribution of *K. griseola* secondary metabolite biosynthetic gene clusters as predicted by antiSMASH where 23 biosynthetic genes clusters were predicted within contigs 1, 2 and 4; natural product classes are colour-coded as follow: terpene (pink), polyketide (blue), nonribosomal peptide (red), polyketide-nonribosomal peptide hybrid (purple), lantipeptide (yellow), siderophore (orange), bacteriocin (green) and unclassified (black).

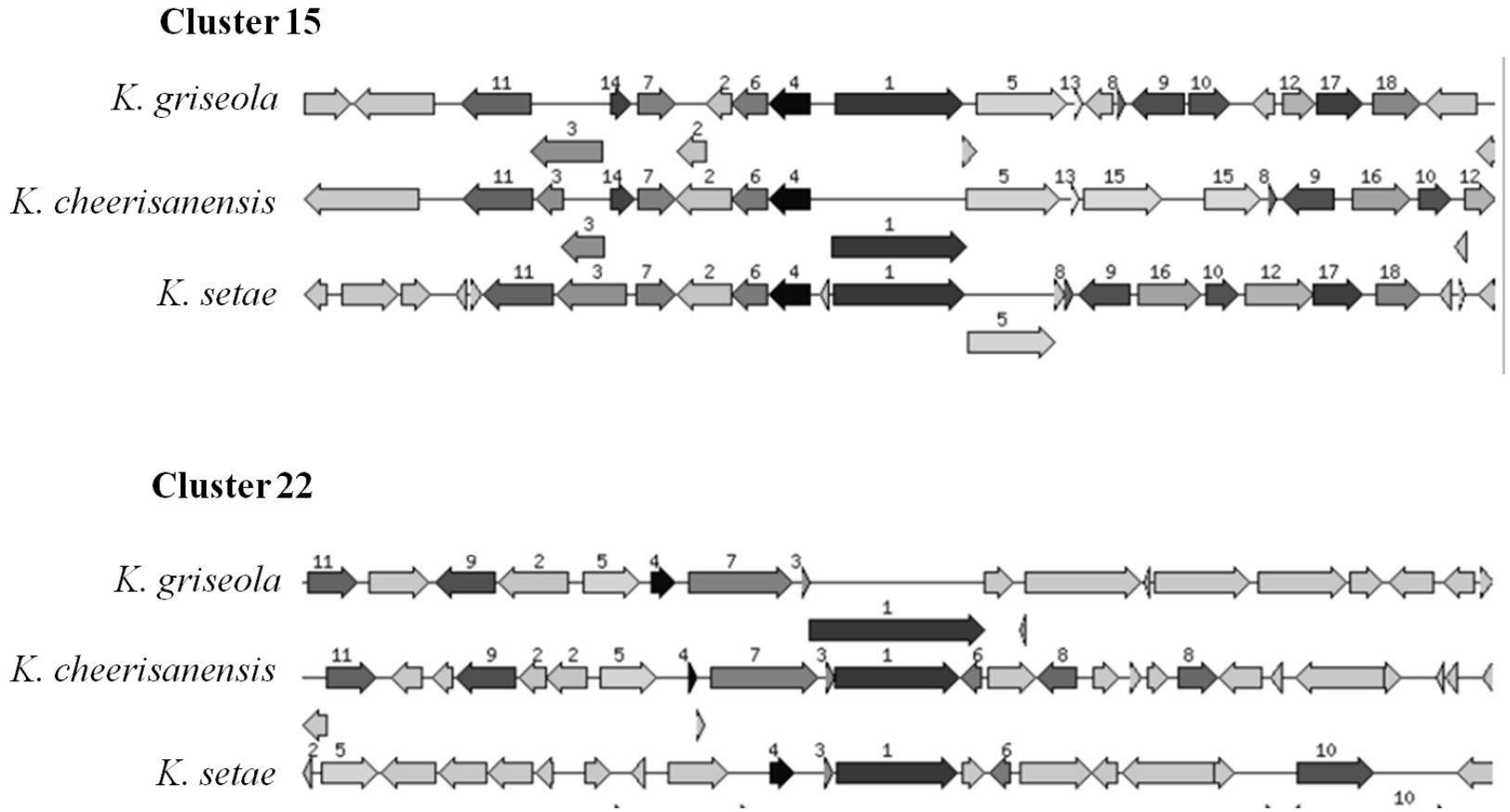


Figure 5.14. Synteny surrounding lantipepide synthetases (labelled as1) illustrating that similar lantibiotic biosynthetic genes to those of cluster 15 and 22 of *K. griseola* were also found in *K. cheerisanensis* and *K. setae*.

as acetyl-, propionyl- and benzoyl-CoA[10]. Six polyketide biosynthetic gene clusters were detected by antiSMASH, three of which had known products.

The bafilomycin biosynthetic gene cluster was identified in *K. griseola* (cluster 21; KGR_431-449) as was expected given the isolation of the compound in previous chapters of this thesis and reports of the compound (also known as setamycin) by *K. griseola* NRRL-16229[86]. Orthologs of this type 1 PK cluster were present in *K. setae* and *K. cheerisanensis*, as well as numerous streptomycetes including those previously reported to produce bafilomycins such as *S. lohii*[77]. A portion of the cluster was found at the end of a contig in *Kitasatospora* sp. MBT66 and was not located in strain MBT63; completing the genomes of these strains would reveal whether all kitasatospora sequenced to date are capable of producing bafilomycins. Given the high incidence of isolating bafilomycins from streptomycetes[28], it is not surprising that kitasatosporae would also have the biosynthetic machinery leading to their production.

The second type 1 polyketide biosynthetic gene cluster with known metabolite was cluster 24 (KGR_1149-1168) and contained genes involved in satosporin biosynthesis as described in Chapter 4,Section 4.3.4. AntiSMASH had categorized cluster 24 as a mixed NRP-PK (type 1) due to the presence of NRPS genes to which a function in satosporin biosynthesis could not be attributed. No orthologous gene cluster could be identified from any other organism. Finally, gene cluster 9 (KGR_3688-3695) contained the type 2 polyketide *whiE* genes involved in the production of the aromatic spore pigment described previously (Section 5.3.2). Orthologous gene clusters were located in all kitasatosporae and numerous streptomycetes, as was expected given the common production of this pigment during sporulation[198].

The remaining three PKs, one NRP and two NRP-PK hybrids had unknown products. Cluster 12 (type 3 PK) had orthologs in all kitasatosporae except *Kitasatospora* sp. MBT66, Cluster 7 had orthologs in *K. cheerisanensis* and the other unknown biosynthetic gene clusters had no orthologs. Further analysis of these unknown biosynthetic gene clusters may provide promising new polyketide and nonribosomal peptide natural products.

Terpenes

Terpenes are structurally diverse, clinically relevant natural products biosynthesized from repeating units of two simple precursors: isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP).

All three *K. griseola* terpene biosynthetic gene clusters had known metabolites. The first, cluster 8, encoded genes for the biosynthesis of the diterpene terpentecin as previously reported from this strain[64, 67]. This included mevalonate pathway (KGR_3233-3239) and terpentecin specific genes (KGR_3241-3247); no orthologs of the terpentecin genes were found in any other organism.

Cluster 11 (KGR_4014-4032) encoded genes for hopanoid biosynthesis (pentacyclic triterpenes) such as the squalene hopene synthase gene (*shc*/*hpnF*), a gene encoding a protein required for the addition of adenosine to the hopane skeleton (*hpnH*) and the aminotransferase gene required for aminobacteriohopanetriol synthesis (*hpnO*)[199]. Orthologous clusters were identified from all kitasatosporae with exact gene synteny and from numerous *Streptomyces* which shared similar clusters.

The last terpene cluster identified by antiSMASH, cluster 19, included the germacradienol-geosmin synthase gene (KGR_6572) responsible for germacradienol and geosmin biosynthesis from farnesyl diphosphate[200]. Geosmin is the main compound, produced by numerous *Streptomyces* and other actinomycetes, which gives soil its earthy smell[201]. Orthologs to KGR_6572 were identified from all kitasatosporae, though synteny of adjacent genes was not always observed.

Siderophores

Siderophores are small molecules responsible for iron uptake in bacteria, fungi and graminaceous plants due to their high affinity for iron[202]. Genes encoding for both the biosynthesis of the compound and transport of the (siderophore+iron) complex are usually

present. Although siderophores do not usually display antimicrobial properties, prodrugs can be designed to take advantage of the siderophore iron uptake system as an entry route[202].

The siderophore biosynthetic genes of cluster 1 (KSE_1692-1695) contained two siderophore synthetases (Type A and C) with similarity to IucA_IucC and RhbC family of proteins, a siderophore synthase with similarity to AlcB (*N*-acetyltransferase) and a lysine/ornithine *N*-monoxygenase gene. Although the biosynthesis of many siderophores involves a protein of the nonribosomal synthetase family to form a polypeptide, this siderophore involved a nonribosomal synthetase-independent pathway that is also found in siderophores such as aerobactin (*iucABCD*), rhizobactin (*rhbABCDEF*), desferrioxamines (*desABCD*), alcaligin (*alcABCDE*), and others. A similar pathway has now been shown to be widely distributed in over 40 species of bacteria[203]. Orthologs of this gene cluster were located in all kitasatosporae as well as in *S. coelicolor* (SCO5799-SCO5801) and other streptomycetes with conserved gene synteny (Figure 5.15). Although the product of cluster 1 remains unknown, it is clear that its production is widespread amongst both streptomycetes and kitasatosporae.

The second gene cluster predicted encoding a siderophore was cluster 14 (KSE_4489-4498). Similarly, the biosynthesis of this unknown compound does not involve a nonribosomal synthethase, but rather a siderophore synthetase (type B), three other biosynthetic genes and 5 transport genes. Orthologous clusters existed in all kitasatosporae except *Kitasatospora* sp. MBT63 (Figure 5.15(B)) and did not seem to be present in any *Streptomyces*, though antiSMASH and SEED predicted orthologous clusters in *Renibacterium salmonarium*, *Polaromonas* sp. JS666, *Streptosporangium roseum*, etc. As previously stated for *K. setae*, no desferrioxamine biosynthetic genes were located in *K. griseola* or any of the other kitasatosporae even though they are often present in streptomycetes[136].

Bacteriocin and unclassified natural products

Bacteriocins are ribosomally synthesized antibacterial peptides produced by bacteria that

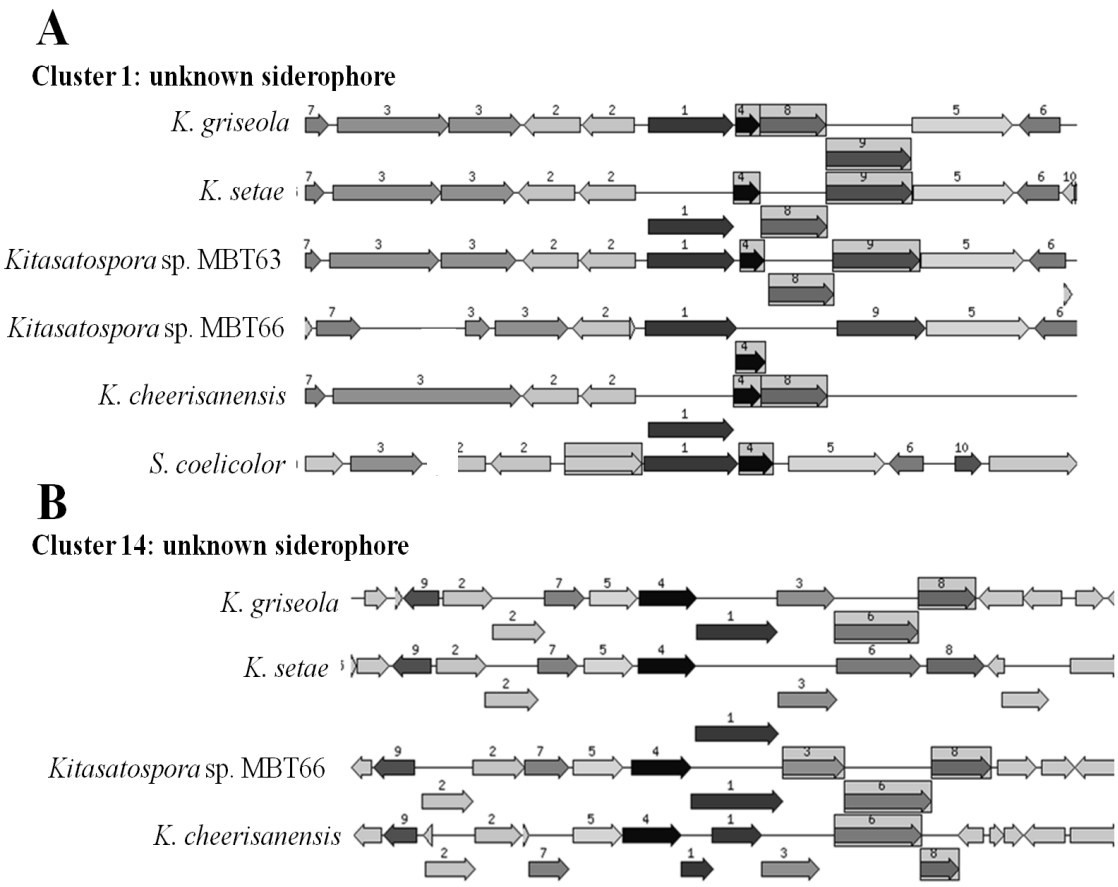


Figure 5.15. Unknown siderophore biosynthetic gene clusters 1 **(A)** and 14 **(B)** in *K. griseola* were found to have orthologous clusters in numerous kitasatosporae and in the case of cluster 1, orthologs were also present in numerous streptomycetes. Orthologous genes are represented by the same number.

are believed to be implicated in colonization, competition and signalling within complex microbial communities[204]. Some have also been shown to be useful pharmaceuticals against pathogenic microorganisms such as methicillin-resistant *Staphylococcus aureus* (MRSA) and *Clostridium difficile*[204]. Cluster 13 (KGR_4156-4164) was predicted by antiSMASH to biosynthesize a bacteriocin that was present in both *K. setae and K. cheerisanensis*. Genes within this cluster had homology to linocin M18 bacteriocin structural gene (KGR_4163), putative membrane protein gene (KGR_4158) and a gene encoding a cytochrome P450 hydroxylase (KGR_4161). AntiSMASH predicted orthologous genes in *K. setae* (KSE_19150-19230), while RAST predicted orthologous genes in *K. cheerisanensis.* It is possible that the bacteriocin produced by *K. griseola* and encoded by KGR_4163 (267 aa) may be similar to linocin M18 which is encoded by *lin* and results in a protein of 266 aa in *Brevibacterium linens* M18[205].

AntiSMASH predicted a compound in the unclassified category of secondary metabolites (KGR_6679-6694) with orthology to KSE_27320-27430, which had been previously identified as genes similar to those involved in valanimycin biosynthesis[136]. BLAST®p revealed protein similarities to *vlmA*, *vlmB* and *vlmH* and orthologous genes were also detected in *Kitasatospora* sp. MBT63. Valanimycin was shown to have antibacterial activity against Gram-negative and Gram-positive bacteria as well as anticancer activities[206].

5.4 CONCLUSION

The analysis of the *K. griseola* MF730-N6 draft genome has provided a wealth of information pertaining to general features such as genome length, tRNA genes and rRNA operons, DNA replication genes and genes involved in cell wall biosynthesis that are shared with the kitasatosporae currently available and with streptomycetes. Kitasatosporae also seem to have orthologs of almost all developmental regulatory genes (*bld* and *whi* genes) usually involved in the complex life cycle of streptomycetes with the exception of *bldB* and *whiJ*. In contrast to

other *Kitasatospora* genome reports, an ortholog of the *ram* cluster, encoding a protein similar to sapB, was located in both *K. griseola* and *K. cheerisanensis* suggesting that some kitasatosporae may use this protein to facilitate aerial hyphae formation while others may not. This research also illustrates the efficiency of the PacBio sequencing technology to provide high quality genome data in few contigs at a reasonable cost.

This research highlights the usefulness of studying kitasatosporae as sources of novel natural products in a drug discovery program as 15 of the 23 secondary metabolite biosynthetic gene clusters predicted by antiSMASH in *K. griseola* had unknown products. This large number of gene clusters supports previous suggestions that non-streptomycete Actinobacteria are a prolific source of novel natural products[207]. Since *Kitasatospora* genomic data consistently exceeds the number of experimentally isolated compounds, it can be concluded that numerous natural products remain to be isolated from members of this underexplored genus. Further analyses into the metabolomes of kitasatosporae may thus provide previously uncharacterized compounds; this will be the subject of the next chapter.

CHAPTER 6 - DISCOVERY OF NEW NATURAL PRODUCTS FROM MEMBERS OF THE GENUS *KITASATOSPORA* BY METABOLOMIC PROFILING

Collaborators and their contributions to this chapter:
Chemical characterizations (Sections 6.2.4, 6.3.2 and 6.3.3) were carried out by Krista Gill, PhD candidate, Molecular and Macromolecular Sciences (MMS), UPEI.

Part of this chapter contains material previously published as:

Gill, K.A., F. Berrue, J.C. Arens, and R.G. Kerr, Isolation and structure elucidation of cystargamide, a lipopeptide from *Kitasatospora cystarginea*. *J. Nat. Prod.*, 2014. **77**(6): 1372-6.

Gill, Krista; Berrue, Fabrice; Arens, Jennifer; Carr, Gavin and R.G. Kerr, Cystargolides, 20S proteasome inhibitors isolated from *Kitasatospora cystarginea*. *J. Nat. Prod.,* 2015. 78(4): 822-6

6.1 INTRODUCTION

Successful drug discovery programs now require a more strategic plan than simply fermenting as many microbes as possible and, ideally, involve innovative sampling and cultivation techniques to uncover underexplored taxa, the development of new assays targeting drugable biological targets and the use of automated analytic techniques to rapidly identify new compounds[19, 28]. This dereplication step is essential to avoid wasting time and financial resources towards the isolation and structure elucidation of known compounds[19]. In a study carried out by the Fundación Medina in Spain, 42% of the 5,530 streptomycetes fermented produced known compounds, whereas only 19% of the 3,140 non-*Streptomyces* actinomycetes produced known compounds[28], illustrating the necessity for both early compound dereplication and selective isolation for the investigation of underexplored genera.

The objective of this study was to use the metabolomic analysis method developed in our research group by Forner et al.[24] to rapidly identify putatively new natural products from a large set of LC-HRMS data from crude extracts of numerous members of non-*Streptomyces* actinomycetes of the genus *Kitasatospora*.

6.2 MATERIALS AND METHODS

6.2.1 Fermentation and extraction of 12 kitasatosporae

Kitasatospora griseola MF730-N6 was obtained from the International Patent Organism Depositary in Japan and eleven strains labelled as *Kitasatospora* spp. were obtained from the ARS culture collection (Table 3.1). Following growth on solid lean medium (Table 2.3), two seeds were grown in 10 mL seed medium (Table 2.3) in 25 X 150 mm glass culture tubes for 48 h each at 30°C and 200 rpm using a scraping from the agar plate for seed 1 and a 7.5% inoculum of seed 1 for seed 2. Each organism was then grown in 10 mL of both lean and ISP2 media (see Table 2.3) in triplicate culture tubes, each inoculated using 500 µL of seed 2 and grown for 48 h at 30°C and 200 rpm. Media blanks were included as negative controls.

All 78 culture tubes were then extracted by adding an equal volume (10 mL) EtOAc directly to the fermentation, allowing 1 h for the extraction at 200 rpm, removing the organic layer and repeating the process a second time. The EtOAc layers were then combined and dried under air and analysed by LC-HRMS (0.5 mg/mL in MeOH) over three runs (26 samples per run) using Thermo Scientific™ LTQ Exactive equipment with a Core Shell Kinetex 1.7 µm C_{18}, 50 X 2.1 mm, 100 Å column, photodiode array detector, SEDEX 80 Sedere LT-ELSD and a Finnigan LXQ ion trap mass spectrometer (MS) in positive mode equipped with an electrospray ionization source (ESI) detecting *m*/*z* between 190 and 2000. Chromatographic separation was achieved with a 4.8 min gradient of 95:5 water:ACN to 100% ACN, 3.2 min of 100% ACN and finally a return to starting conditions over 3 min; solvents contained 0.1% formic acid and a flow rate of 500 µL/min was used. Methanol blanks were included as negative controls and reserpine (1 µg/mL final concentration) was added to each sample as an internal control.

6.2.2 Data processing and statistical analysis

High resolution mass spectral data for all samples and methanol blanks were processed as described by Forner et al.[24] with a few modifications. Briefly, the .RAW files generated by the instrument were first converted to .CDF files using the Xcalibur™ software to enable import into MZmine. MZmine 2[208] was then used to carry out mass detection, chromatogram building, de-isotoping, alignment and export to Excel® as described previously[24]. Peak intensity threshold during mass detection was set to 1E4 while mass to charge ratio (*m*/*z*) and retention rime (t_R) tolerances during bucket generation were set to 0.005 and 0.1 min, respectively; other parameters were set to default.

Within Excel®, replicate buckets were first condensed and those also present in the solvent (MeOH) and media samples were removed by selecting and deleting buckets that were present in the appropriate media and any of the methanol blanks. This process was performed separately for each medium due to their different compositions. Unlike the method of Forner et al.[24], the area under the curve for each bucket was not converted to a binary format. Statistical analysis

was then performed using the principal component analysis tool within the program The Unscrambler® X (Camo, Woodbridge, USA) using crude extract as the samples and buckets as the variables.

6.2.3 Biological assay

Crude extracts were tested following the clinical laboratory standards institute testing standards (2003) in a microplate broth assay, at a concentration of 75 µg/mL, against the following microorganisms: methicillin-resistant *Staphylococcus aureus* ATCC 33591, vancomycin-resistant *Enterococcus faecalis* EF 379, *Candida albicans* ATCC 14035, *Pseudomonas aeruginosa* ATCC 14210, *Proteus vulgaris* ATCC 12454 and *Staphylococcus warneri* ATCC 17917. Media blanks and solvent containing 1 µg/mL reserpine were also tested.

6.2.4 Purification of compounds of interest and structure elucidation

To follow up hits, the appropriate organism was re-fermented as mentioned previously except using Fernbach flasks containing each 500 to 750 mL of the appropriate medium. The fermentation was extracted as described and the resulting crude extracts were then given to a PhD candidate in our laboratory (Krista Gill, Molecular and Macromolecular Sciences) in order to purify the desired compound using various chromatographic separation techniques and to carry out the structure elucidation using one- and two-dimensional nuclear magnetic resonance (NMR) experiments and (LC-HRMS).

6.3 RESULTS AND DISCUSSION

6.3.1 General results

Following LC-HRMS analysis of crude extracts from 12 kitasatosporae grown in two media, the 81 files generated by the instrument were used for the metabolomic analysis. This included data from kitasatosporae fermentations, media blanks and methanol samples.

To analyse this large amount of data, the metabolomic analysis method described by Forner et al.[24] was used. The data was processed using MZmine 2 which first generated a list of all masses present at each scan of each data file above the threshold chosen (1E4). This threshold was based on inspecting methanol blank mass spectra and choosing the lowest intensity above

most background peaks and also depended on the computational capabilities of the computer used for the analysis. The list of exact masses generated was then used by the chromatogram builder module to detect the same masses (within 0.005 atomic mass units) in adjacent scans; only masses which were present in all scans within a 0.1 min period were used to build the chromatogram. De-isotoping was then performed to keep only the most intense isotope. Finally the Alignment module allowed the formation of the buckets, combining the *m*/*z* and retention time values for each peak in the chromatograms and aligning all of the buckets throughout all samples into one document that was then exported into Excel®.

Inspection of this Excel® file revealed that the internal standard, reserpine (predicted pseudomolecular ion $[M+H]^+$ *m*/*z* of 609.2812), which had been added to every sample was detected as a single bucket (*m*/*z* 609.2812, t_R 3.11 min) in all samples (excluding the MeOH blanks ran by technician) and confirmed there was no shift in mass detection or retention time between the three LC-HRMS analyses conducted over eight days.

From the crude extract of 12 kitasatosporae fermented in two media and in triplicate, MZmine 2 generated a total of 633 buckets. Visual inspection of the Excel® file revealed that numerous buckets were sometimes generated for the same *m*/*z* with retention times throughout the analysis (Figure 6.1 (A)). Condensing these replicates into a single bucket was essential for the proper removal of media components and artefacts and resulted in 527 buckets.

A total of 34 buckets were present in at least one of the MeOH blanks and were thus removed from the list. Fifty-seven buckets were detected from at least one the the lean medium control fermentation extracts and those peak areas were thus replaced by zero for fermentations carried out in lean medium. Similarly, data from 39 buckets were "zeroed" in samples from cultures grown in ISP2 due to their presence in the media blanks. Artefact and media component suppresion resulted in a total of 444 buckets which represented ions of compounds produced as a result of bacterial metabolism.

A

Samples

Buckets (*m/z*, retention time)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	row I	row m/z	row retention time		JC11_97_I	JC10_77_2	JC12_25_I	JC13_69_2	JC15_74_2	JC14_19_I	JC16_57_I	JC17_52_I	JC19_83_I
39	544	213.1462	1.04		0	0	0	0	0	0	0	0	0
40	9	213.1462	0.66		884054.2	490526.8	812755.5	735958.8	630613.6	505265.3	600110.3	706499.5	747328.6
41	287	213.1463	1.23		0	0	0	0	0	0	0	0	0
42	631	213.1463	1.94		0	0	0	0	0	0	0	0	0
43	56	213.1599	2.15		0	0	175066.9	135378.8	95869.47	0	154634.2	0	0
44	559	217.1337	2.76		0	0	0	0	0	0	0	0	0
45	51	219.0541	0.90		0	0	2.37E+07	2.45E+07	0	2.41E+07	0	2.42E+07	2.40E+07
46	1	219.0541	0.77		2.26E+07	2.38E+07	0	0	0	0	0	0	0
47	629	219.0541	1.68		0	0	0	0	0	0	0	0	0
48	80	219.0541	1.06		0	0	0	0	2.37E+07	0	2.37E+07	0	0
49	329	219.0541	1.33		0	0	0	0	0	0	0	0	0
50	379	219.0541	1.19		0	0	0	0	0	0	0	0	0
51	373	219.0541	1.40		0	0	0	0	0	0	0	0	0
52	395	219.0541	1.83		0	0	0	0	0	0	0	0	0
53	168	219.0976	2.34		0	0	0	0	0	0	0	0	0
54	516	219.113	1.73		0	0	0	0	0	0	0	0	0
55	192	219.1162	2.34		0	0	0	0	0	0	0	0	0

B

Buckets (*m/z*, retention time)

72 samples (12 species, 2 media, triplicate)

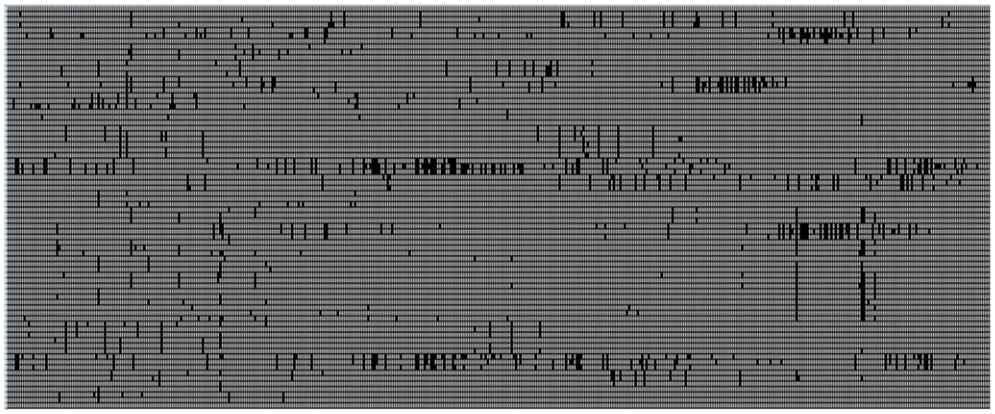


Figure 6.1. Example of buckets (highlighted in grey) which required being condensed into one bucket in order to properly carry out the artefact suppression step **(A)**. Matrix generated during metabolomic analysis of LC-HRMS data collected from the crude extracts of 12 kitasatosporae cultured in two media and in triplicate (rows) giving rise to 444 different buckets (columns) illustrating the overwhelming amount of information generated during this experiment. Black boxes represent the presence of that bucket in the sample **(B)**.

The matrix generated (Figure 6.1 (B)) contained 72 rows, each representing a single crude extract, and 444 columns, each representing a single bucket containing information pertaining to both the *m*/*z* and t_R of the ions detected by the mass spectrometer. Each black square represents the presence of the particular bucket within the sample. Although it is possible to draw a few general conclusions such as some crude extracts contained more metabolites than others and some metabolites were produced by numerous bacteria while others appeared to be produced by a single organism, it was not possible to analyse this overwhelming amount of data set in any detail.

The principal component analysis allowed the rapid visualization of the metabolomes which differed most within the dataset. *Kitasatospora cystarginea* (NRRL B-16505) and *K. mediocidica* (NRRL B-16109) had the most unique metabolomes (Figure 6.2) as their metabolites, represented here as buckets, contributed to most of the variation within the data set (56% and 17%, respectively). The loadings plot (Figure 6.3) indicated that the variable (bucket) causing *K. mediocidica* to separate from the remainder of the data had a *m*/*z* 686.4711(t_R 3.08), while ions with *m*/*z* 672.4557 and 708.4526 also contributed, but to a lesser extent. Similarly, the buckets causing *K. cystarginea* to separate from the remainder of the data had *m*/*z* 371.2177 (t_R 3.09) and 357.2022 (t_R 2.88).

The strength of this analysis lies in its rapid detection of metabolites that are unique to a particular organism, thus increasing the chances that the compounds have not previously been described. Furthermore, this technique provides information on the potential novelty of the compound based on a rapid database search in AntiBase 2012 using the *m*/*z*. The use of chromatographic separation (LC) in combination with UV, MS, MS/MS or NMR is regarded as a very sensitive and powerful dereplication tool for the identification of new natural products[23]. In contrast, a disadvantage of this technique is that it does not provide any information as to the potential biological activity of the compound in treating various human and animal diseases.

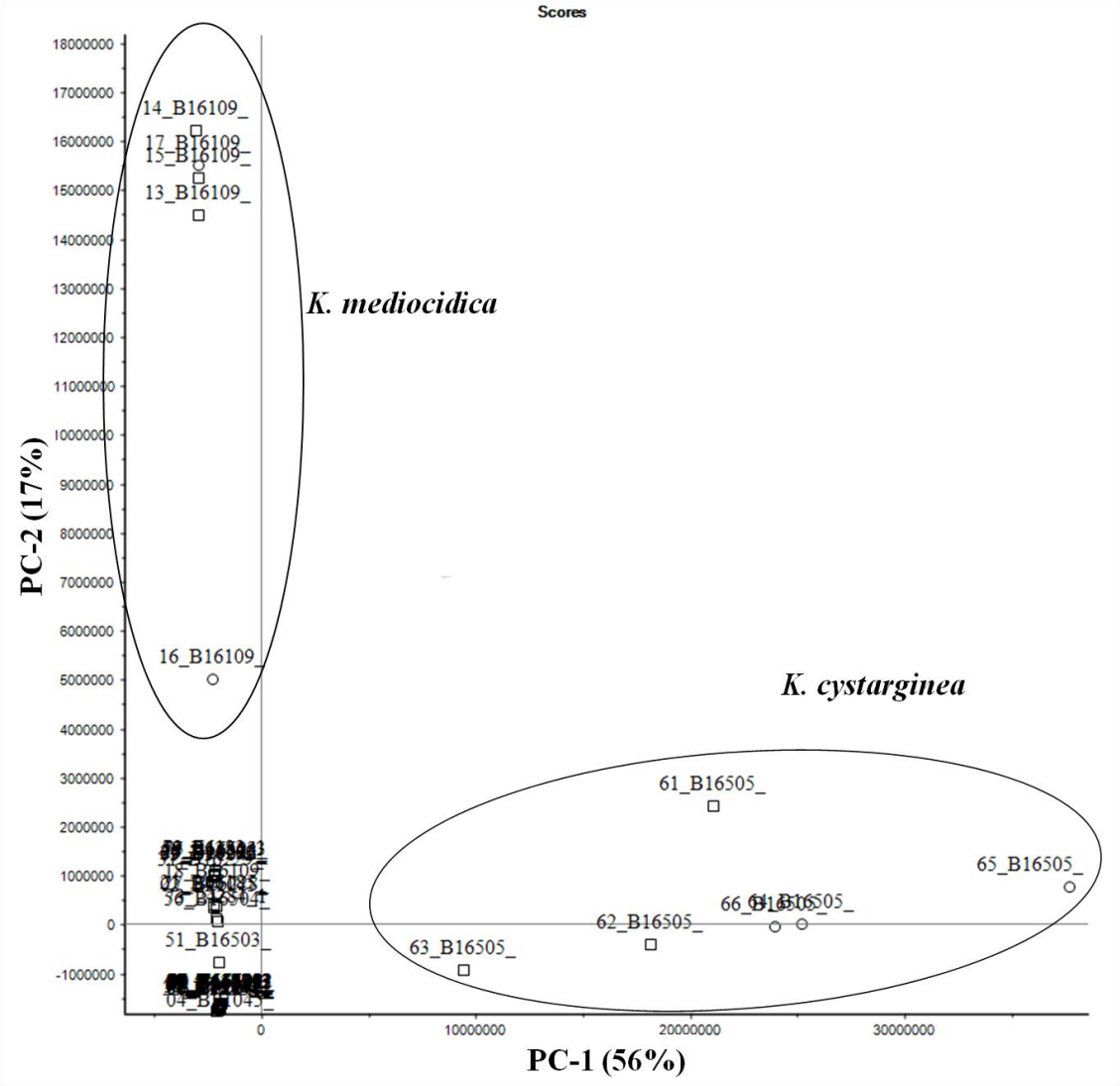


Figure 6.2. Scores plot of principal components 1 and 2 illustrating that the metabolomes of *K. mediocidica* (B-16109) and *K. cystarginea* (B-16505) were the largest contributors of variability within this data set. Circles represent extracts obtained from lean medium fermentations while squares represent extracts obtained from ISP2 medium fermentations.

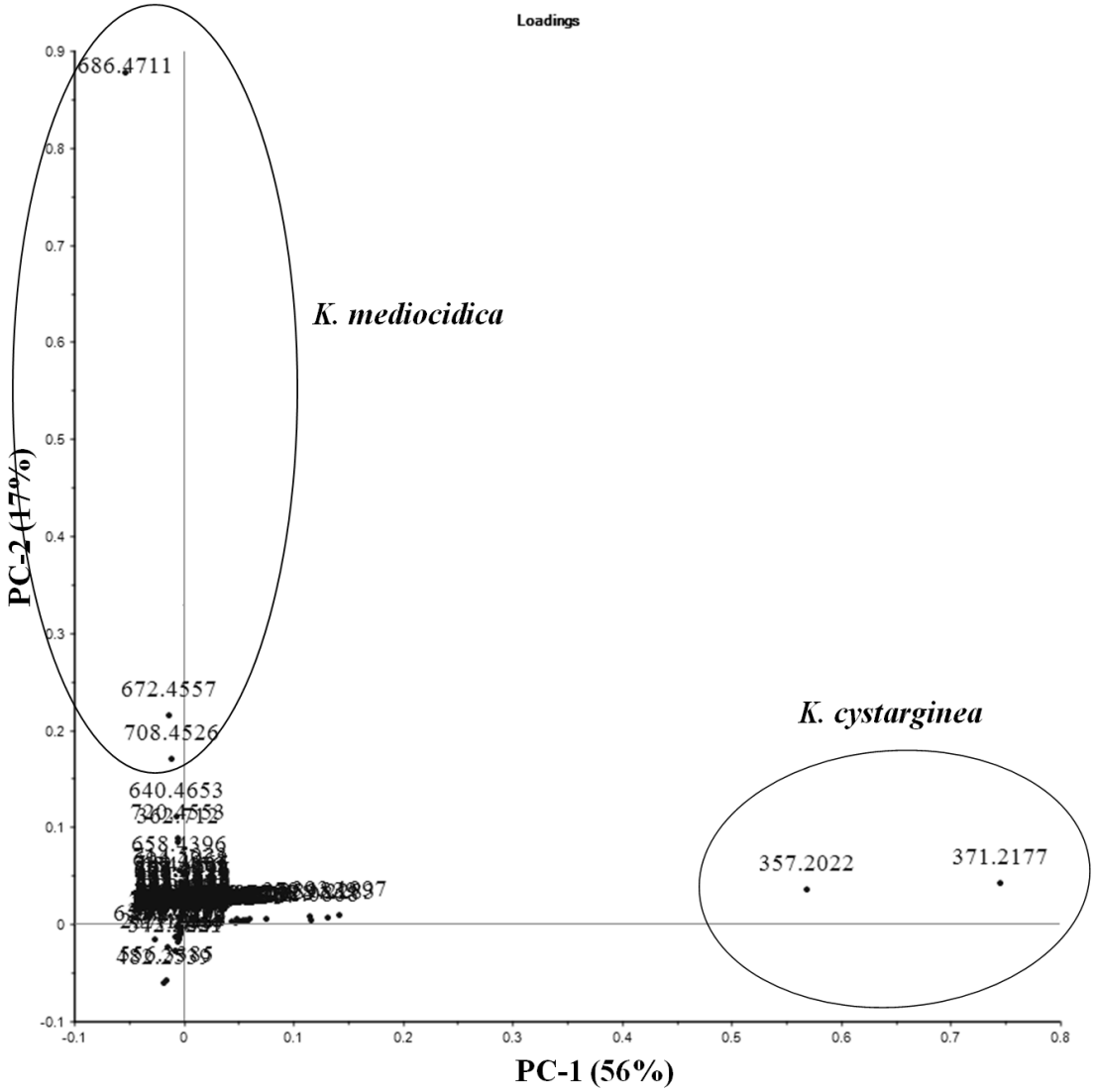


Figure 6.3. Loadings plot of principal components 1 and 2 illustrating which buckets (represented by the *m*/*z*) were causing *K. mediocidica* and *K. cystarginea* to separate from the remainder of the data.

As one of the objectives in the laboratory is to find novel secondary metabolites that could be developed into human or animal pharmaceutical products, knowledge of the activity of the crude extracts against numerous microorganisms can aid in prioritizing compound purification and structure elucidation. Of the 79 samples submitted for antibacterial assay (sample concentration = 75 µg/mL) against methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant *Enterococcus faecalis* (VRE), *Candida albicans, Pseudomonas aeruginosa, Proteus vulgaris* and *Staphylococcus warneri,* crude extracts from three organisms showed moderate or good antimicrobial activity. All three replicate extracts of *K. kifunensis* NRRL B-24284 grown in lean medium showed 89, 92 and 95% growth inhibition of *C. albicans,* while all three replicates of *K. streptosporus* NRRL B-16228 grown in ISP2 showed 62, 79 and 83% growth inhibition of VRE. Finally, replicates 1 and 2 of *K. griseola* NRRL B-16229 grown in lean medium showed 94% growth inhibition of *S. warneri* and replicate 2 also showed 72% growth inhibition of MRSA.

Since there appeared to be only minute quantities of the bioactive compound in the extracts of *K. streptosporus* (the ELSD chromatograms were devoid of any significant peaks), this extract was not investigated further as the scale necessary to purify and elucidate the structure of the compound exceeded the time and resources available at this point. Other samples highlighted by the principal component analysis or by the antibacterial assay were investigated further by Krista Gill (PhD candidate MMS).

6.3.2 Hits based on chemical novelty

Ions with *m*/*z* 371.2177 and 357.2022 from *K. cystarginea* extracts

The two variables (buckets) contributing to most of the variability in the data set were those with *m*/*z* 371.2177 (t_R 3.09 min) and 357.2022 (t_R 2.88 min) in extracts of *K. cystarginea* grown in both ISP2 and lean media. The only compound previously reported by *K. cystarginea* was cystargin ($C_{20}H_{42}N_8O_6S$, $[M+H]^+$ 523.3021)[209], suggesting that neither of the highlighted ions was cystargin. Searches for known compounds within AntiBase with $[M+H]^+$ mass within 5

ppm of both *m/z* did not generate any hits, though numerous compounds had masses consistent with $[M+Na]^+$; all had molecular formula $C_{21}H_{32}O_4$ and $C_{20}H_{30}O_4$ for *m/z* 371.2177 (14 hits) and 357.2022 (47 hits), respectively. Since, the sodium adducts were also present within the same mass spectra ($[M+Na]^+$ *m/z* 379.1839 and 393.1997), it could be concluded that the ions initially observed belonged to the pseudomolecular ion $[M+H]^+$ and, thus, were not present within AntiBase. Given the lack of hits and the fact that these two compounds seemed to be the major constituents of the crude extracts from *K. cystarginea* (Figure 6.4 (A)), crude extract from a 5 L fermentation was given to Krista Gill for purification and structure elucidation.

She was able to elucidate the structure of the compounds with ions of *m/z* 357 and 371 and confirmed that they corresponded to two novel depsipeptides, cystargolide A and B, with molecular formula $C_{17}H_{28}N_2O_6$ and $C_{18}H_{28}N_2O_6$, respectively; their structures were recently published in the Journal of Natural Products[210]. The purification procedure also allowed her to identify that the compound at 3.60 min of the ELSD (Figure 6.4 (A)) was yet another unknown secondary metabolite produced by *K. cystarginea* with molecular formula $C_{49}H_{59}N_7O_{13}$ and pseudomolecular ion $[M+H]^+$ *m/z* of 954.4287. The structure elucidation paper of this new lipopeptide, cystargamide, was published in the Journal of Natural Products[211] and its structure is illustrated in Figure 6.4 (C). The metabolomic analysis was, thus, very effective at determining ions which were both unique and abundant.

Ions with *m/z* 686.4711, 672.4557 and 708.4526 from *K. mediocidica* extracts

The main variable (bucket) responsible for pulling *K. mediocidica* away from the remainder of the data set had an *m/z* 686.4711 and t_R 3.07 min and constituted a major portion of the crude extract (Figure 6.5 (A)). This observed ion represented the pseudomolecular ion $[M+H]^+$ of the compound since the sodium adduct (*m/z* 708.4526) was also detected at 3.07 min. AntiBase revealed a single hit consistent with this mass, the peptide pepstatin A (produced by *Streptomyces* sp.[212]) with molecular formula $C_{34}H_{63}O_9N_5$. The ion with *m/z* 672.4557 (t_R 2.94 min) was consistent with an analogue of the main compound containing one less methylene

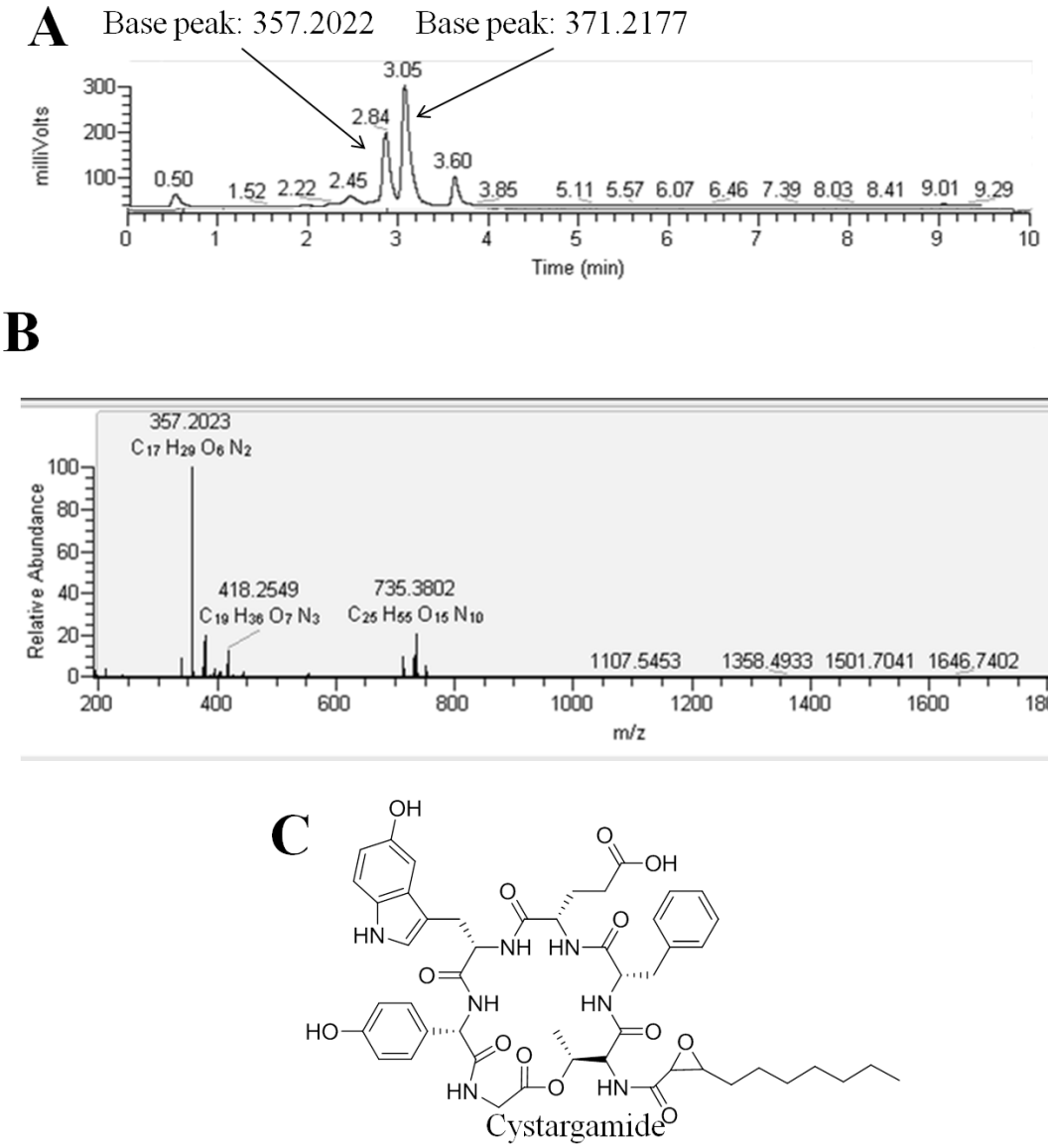


Figure 6.4. Evaporative light scattering detector chromatogram **(A)** and mass spectrum at 2.88 min **(B)** illustrating that compounds with *m*/*z* 371.2177 (t_R 3.09) and 357.2022 (t_R 2.88) constituted the major compounds within the crude extracts of *K. cystarginea.* Structure of a third new secondary metabolite from *K. cystarginea* extract, the new lipopeptide cystargamide **(C)**.

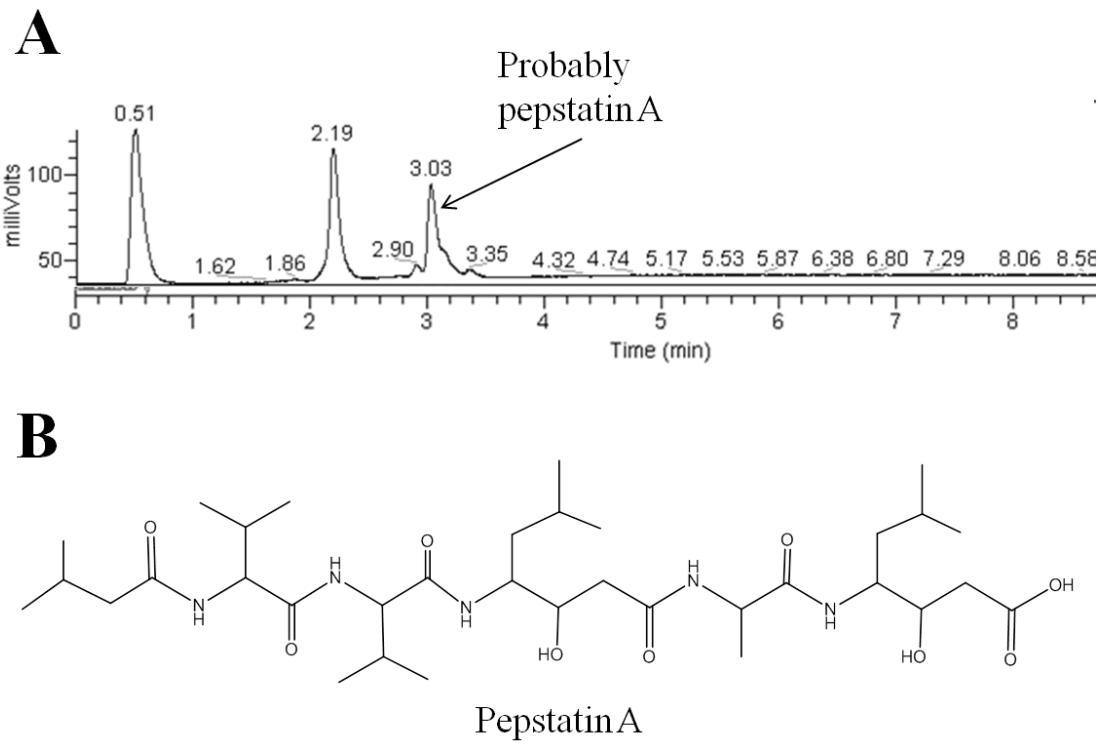


Figure 6.5. Evaporative light scattering detector chromatogram from the crude extract of *K. mediocidica* NNRL B-16109 where pepstatin A was identified as the likely compound detected at 3.03 min **(A)** and structure of the peptide pepstatin A **(B)**.

group (-CH_2-) and had two AntiBase hits: pepstatin BU and pepsidin A, both with molecular formula $C_{33}H_{61}O_9N_5$. Given the high production of the compound in *K. mediocidica* and the few possible hits for the detected ions, the samples were given to Krista Gill for further analysis.

NMR experiments revealed signals with characteristic peptide chemical shifts such as the presence of signals around 7.5 ppm (consistent with the amide protons) as well as methine and methylene signals at approximately 4 and 3 ppm, respectively (consistent with branched aliphatic amino acid side chains), all of which was consistent with the compound pepstatin. Given the high probability that the compound with *m*/*z* 686.4711 was the known peptide pepstatin A (Figure 6.5 (B)), the extract was not investigated further.

A limitation of this technique is its dependence on compound ionization within the mass spectrometer, where potentially novel compounds may not be detected if they do not ionize well. Using this technique in combination with others which do not rely on compound ionization, such as UV, NMR and biological assays, could result in a more comprehensive understanding of the metabolomes under study.

Since more time and resources are necessary to isolate and elucidate the structure of minor components of the crude extracts, identifying unique and highly abundant compounds first was advantageous. However, since novel chemical entities may be present as minor components of the crude extracts, the data can be easily converted to a binary format in order to search through numerous principal components and identify putatively novel compounds.

6.3.3 Hits based on biological activity

Kitasatospora kifunensis NRRL B-24284 extract is active against *Candida albicans*

Since the crude extracts of *K. kifunensis* NRRL B-24284 grown in lean media exhibited strong inhibitory activities against *C. albicans*, it was re-fermented in 750 mL lean medium, extracted as mentioned previously and given to Krista Gill for purification and structure elucidation of the bioactive compound.

NMR analyses revealed that the structure of the compound detected at 3.85 min (Figure 6.6 (A)) was the known compound naphthomycin A, an ansamycin antibiotic. Given that naphthomycin A was reported to have antifungal activity[213], the growth inhibition observed in *C. albicans* was probably a result of the presence of this compound in the crude extracts of *K. kifunensis* grown in lean medium.

Furthermore, the metabolomic analysis revealed that extracts from the lean medium contained ions with *m*/*z* 720.2948 and 742.2769, consistent with the pseudomolecular ion and sodium adduct, respectively, of that expected for naphthomycin A, that were also absent from fermentations of the same organism grown in ISP2 medium. Unfortunately, the ELSD chromatograms of the crude extracts originating from ISP2 medium suggest that very few metabolites were produced in this medium. It is thus possible that another compound was responsible for the antifungal activity. Talosin A and B are compounds known to have inhibitory activity against *C. albicans* (MIC 15 and 7 µg/mL, respectively) that were previously isolated from another strain of the same species[214], though no ions with *m*/*z* consistent with the predicted talosin pseudomolecular ions (417.1180 and 563.1759) or the sodium adducts (439.1000 and 585.1579) were detected in any of the crude extracts. Given the high probability that the antifungal activity of these extracts was caused by a known compound, this extract was not investigated further for novel metabolites.

<u>*Kitasatospora griseola* NRRL B-16229 extract is active against *S. warneri*</u>

Although this organism was known to produce setamycin (bafilomycin B1), which is active against *S. aureus* (MIC 25 µg/mL)[90], ions with *m*/*z* consistent with the pseudomolecular ion ($[M+H]^+$ 816.4528) and the sodium adduct ($[M+Na]^+$ 838.4383) were not detected in the extracts. Compounds were, however, present in the ELSD chromatograms at ~5.5 min (setamycin was previously determined to be detected by the ELSD at 5.50 min in Section 2.3.4), but did not ionize well in the mass spectrometer. It is thus difficult to assess whether the inhibitory activity against *S. warneri* and MRSA of the extract were caused by the presence of

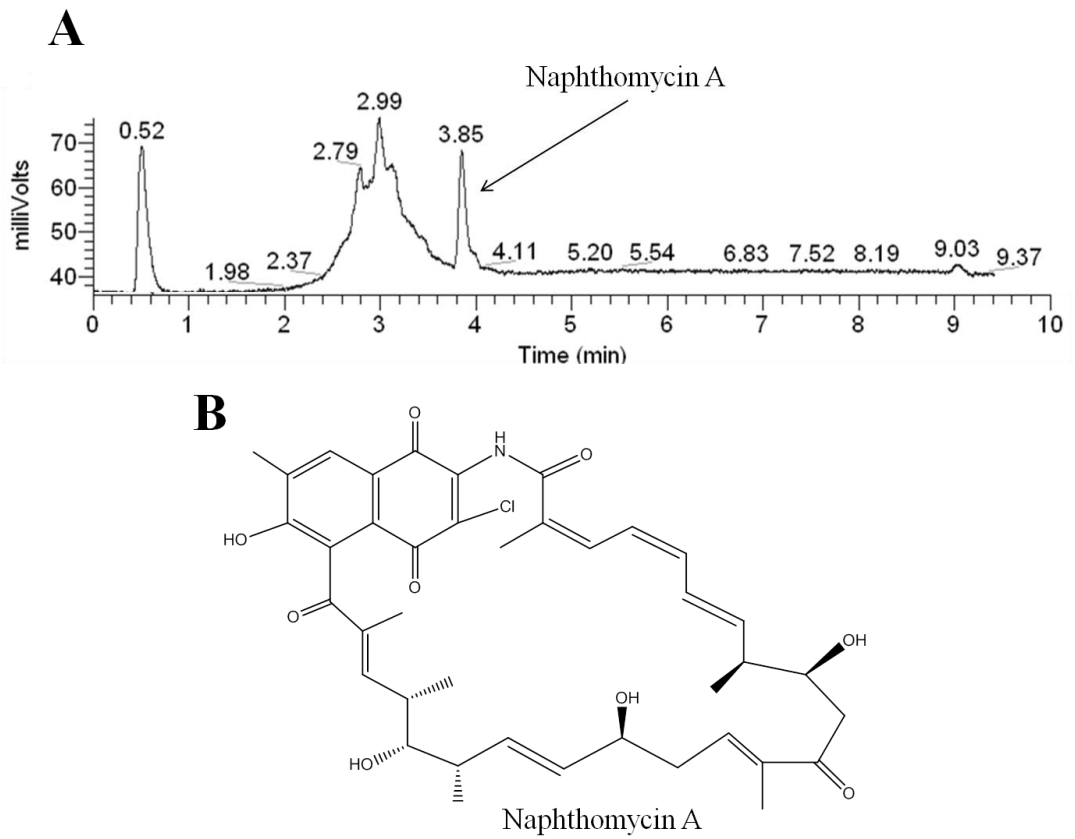


Figure 6.6. Evaporative light scattering detector chromatogram of the crude extract of *K. kifunensis* NRRL B-24284 grown in lean medium **(A)** illustrating that the structure of the compound detected at 3.85 min was the known compound naphthomycin A **(B)**.

setamycin or not. It is also possible that another compound within the extract was responsible for the antibacterial activity. In fact, extracts from *K. griseola* NRRL B-16229 grown in lean medium contained major compounds with *m*/*z* 542.3229 (t_R 3.79) and 556.3389 (t_R 3.93). These were also highlighted by the principal component 3 as buckets responsible for pulling *K. griseola* extracts from the remainder of the data set (Figure 6.7). Since the sodium adducts were also identified (*m*/*z* 564.3047 (t_R 3.79) and 578.3203 (t_R 3.93)), an AntiBase search using the mass of the pseudomolecular ions $[M+H]^+$ revealed that no compounds within the database shared the same molecular formula. Furthermore, these compounds appeared to be produced mostly in cultures from lean medium as the ELSD chromatogram from ISP2 extracts of the same organism were devoid of any significant peaks at 3.74 and 3.90 min (Figure 6.8). The two new compounds are most likely analogues of each other, differing by one methylene group (CH_2) as the two compounds have mass to charge ratios 14 atomic mass units apart. Krista Gill is currently working on purifying and elucidating the structure of these putatively new compounds. Gill was also able to identify the compound 3-methyl-lumichrome from the same extracts (*m*/*z* 257.1033, t_R 2.60), which also appeared to be produced by *K. griseola* MF730-N6 and *K. setae* (presence of same bucket, data not shown). No activity was reported for this compound in AntiBase but lumichrome is a derivative of riboflavin (vitamin B2) that is commonly secreted by bacteria, plants and algae and has been proposed to be involved in quorum sensing[215] and as a root respiration enhancer in plant associated microbes[216]. Though the bioactivity against staphylococci is probably not caused by the presence of this compound, it is beneficial to associate ELSD chromatogram peaks within the extracts with known compounds. The isolation, structure elucidation and bioactivity screening of the purified compounds with mass spectral ions of *m*/*z* 542 and 556 will reveal whether they constitute novel chemical entities and whether these compounds were responsible for the bioactivity observed in the crude extracts against staphylococci.

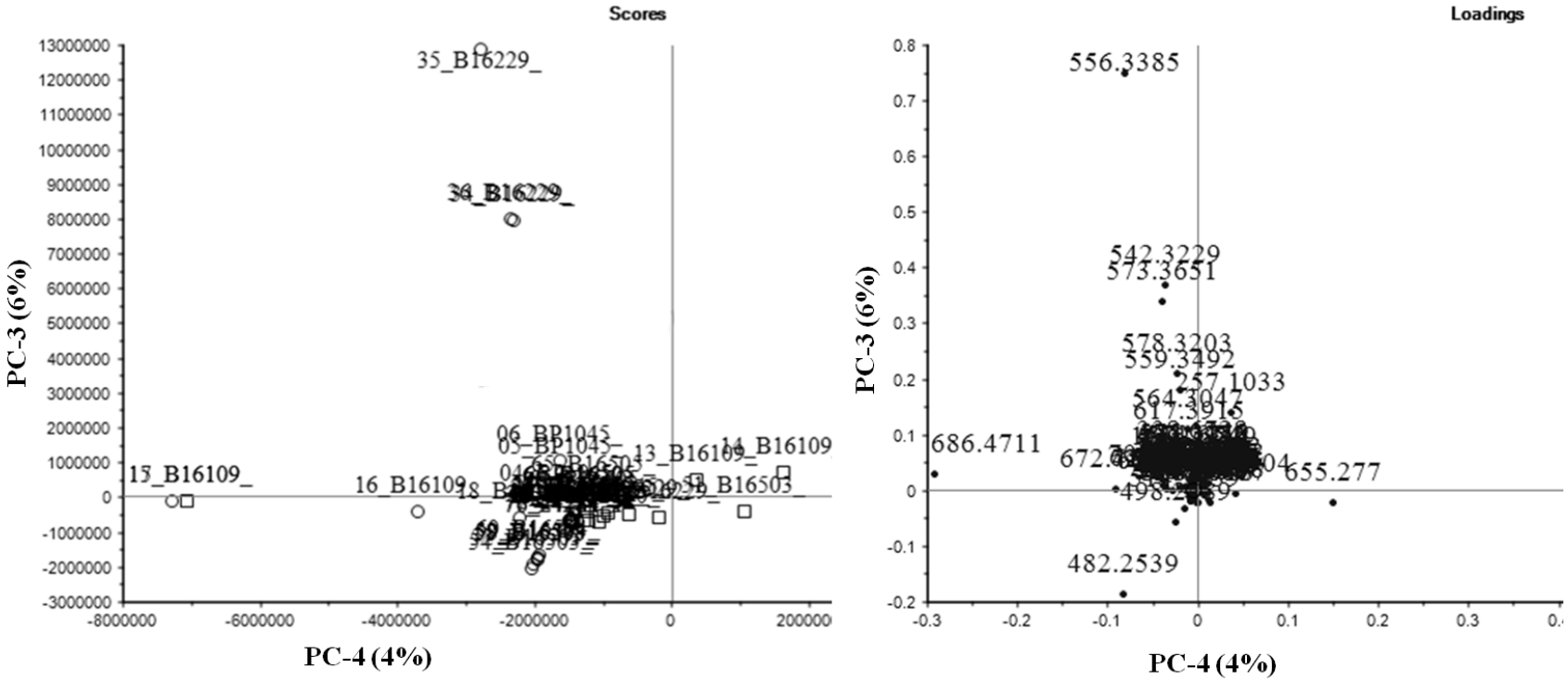


Figure 6.7. Scores and loadings plots of principal components 3 and 4 illustrating that the metabolites responsible for pulling the extracts of *K. griseola* NRRL B-16229 (grown in lean medium) away from the remainder of the data had *m*/*z* 556.3385 and 542.3229. Circles in the scores plot represent extracts obtained from lean medium fermentations and squares represent extracts obtained from ISP2 medium fermentations.

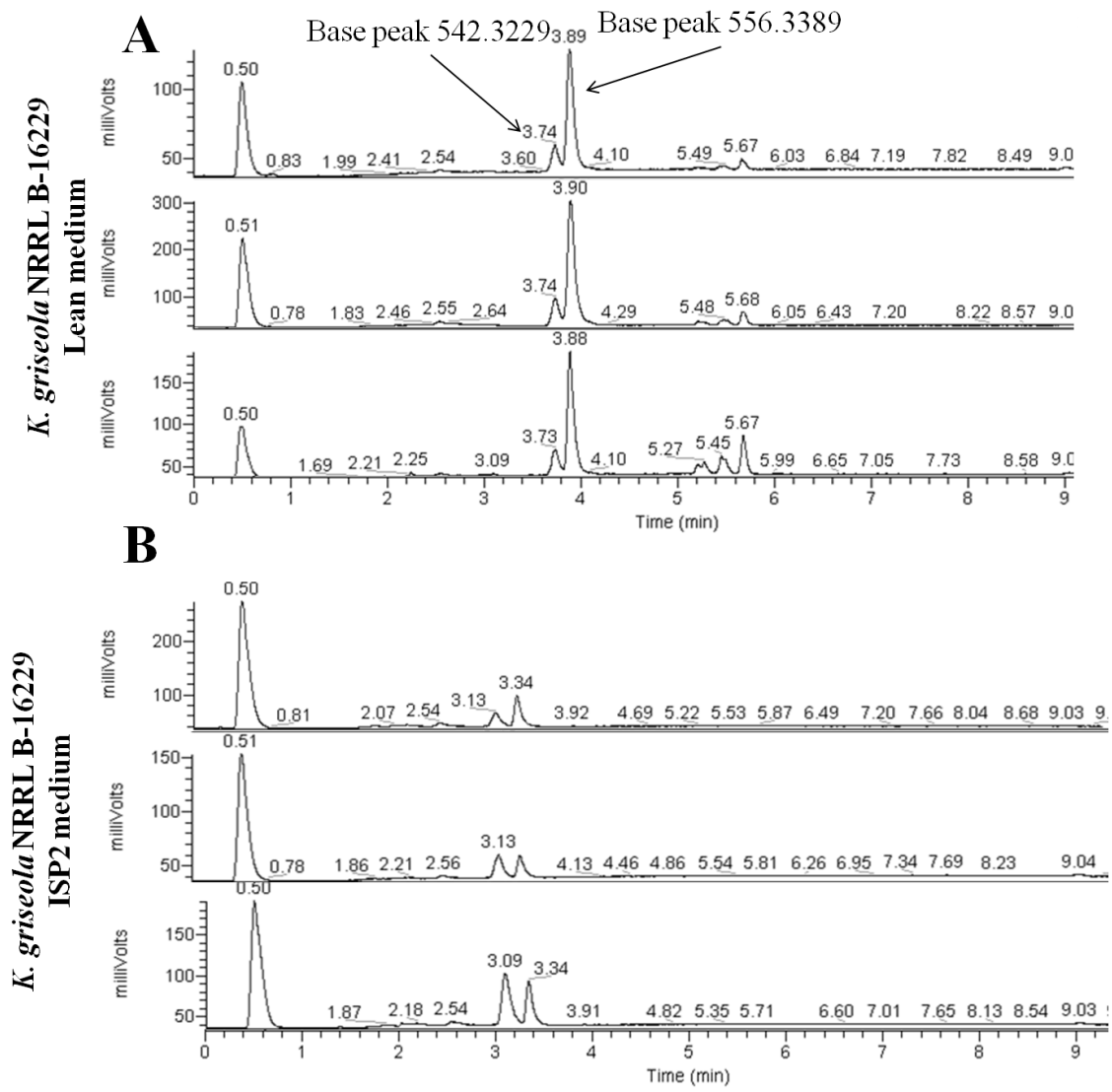


Figure 6.8. Evaporative light scattering detector chromatograms of the crude extracts from triplicate cultures of *K. griseola* NRRL B-16229 grown in lean medium **(A)** and ISP2 medium **(B)**, illustrating the different chemical profile obtained from each medium, where the major compounds produced in lean medium are putatively new compounds with a mass to charge ratio of 542.3229 and 556.3389.

6.4 CONCLUSION

The use of LC-HRMS metabolomics in combination with antimicrobial assays has enabled the discovery of at least three novel natural products, cystargamide and cystargolides A and B from *K. cystarginea* and two putatively new compounds from *K. griseola* (compounds with $[M+H]^+$ *m*/*z* 542 and 556) with possible biological activity against MRSA and *S. warneri.*

Although compound rediscovery was inevitable for both techniques, time and resource investments was considerably lower for hits based on chemical novelty as a simple proton NMR was sometimes able to address the likelihood that the compound was known as was the case for pepstatin A in extracts of *K. mediocidica*. This research supports the use of LC-HRMS based metabolic profiling to rapidly identify new compounds and the use of kitasatosporae as a prolific source of novel natural products in a drug discovery program.

CHAPTER 7 - GENERAL CONCLUSIONS AND FUTURE DIRECTIONS

7.1 Development of a metabolically simplified strain of *Streptomyces lividans* TK24 unable to heterologously produce the antibacterial terpene terpentecin.

In order to provide a practical, renewable and economically relevant source of terpenes, the first objective of this thesis was to develop a bacterial host platform for the heterologous production of bioactive terpenes. *Streptomyces lividans* TK24 was first genetically engineered to eliminate competing pathways leading to the production of actinorhodin and undecylprodigiosin. Genotypic and phenotypic evidence was presented that confirmed *S. lividans* no longer produced ACT and RED.

To generate a recombinant host capable of producing an antibacterial terpene, the terpentecin biosynthetic gene cluster was successfully cloned from *Kitasatospora griseola* MF730-N6 in three fragments and was expressed in *S. lividans* TK24 Δ*act*Δ*red* via a new integrating vector (pJC003) under the control of the constitutive promoter P*ErmE*_actII-ORF4*_JC to drive the transcription of downstream genes. Although genetic studies confirmed that exogenous terpentecin and promoter genes were present in *S. lividans* TK24 ΔactΔred::pTERP1 and RT-PCR confirmed that transcripts of terpentecin gene were present, metabolomic PCA analysis did not suggest any terpentecin production in the heterologous host.

Since terpentecin and antibacterial activity were not detected in the developed heterologous host, the recombinant strain would not, at this moment, provide a rapid method of assessing terpene yields in mutant strains. Future studies attempting to isolate the elusive compound terpentecin from the native producer may provide amounts of pure terpentecin which could be used to develop more sensitive analytical techniques. Alternatively, other antibacterial terpene genes, such as those for phenalinolactone or platencin biosynthesis, could be introduced in the developed host as they have already proven to be produced heterologously in *Streptomyces* hosts[50, 55].

Since the molecular tools generated during this study (pJC003, P*ErmE*_actII-ORF4*_JC and *S. lividans* Δ*act*Δ*red*) were not commercially available, their development during this study

has made them accessible to other members of the Kerr laboratory and to collaborators for future genetic studies involving microorganisms of the genera *Streptomyces* and *Kitasatospora*. As new gene clusters are described, whether from the Kerr laboratory or other research groups, this bacterial host can be used to confirm the products of these pathways and may also act as a reliable source cryptic natural products or compounds in short supply.

7.2 Satosporin A and B are new polyketide natural products from *K. griseola* MF730-N6 with an unprecedented tricyclic ring system and unusual sugar moieties.

In attempting to isolate terpentecin from a large scale (12 L) fermentation of *K. griseola* MF730-N6, two new compounds were detected in the crude extracts. LC-HRMS analysis revealed the presence of ions with *m*/*z* 711.3558 and 549.3021, corresponding to the sodium adduct pseudomolecular ion $[M+Na]^+$ of satosporin A ($C_{34}H_{56}O_{14}$) and B ($C_{28}H_{46}O_9$), respectively. The compounds were purified using liquid partitioning, normal- and reversed-phase silica-based chromatography and HPLC, and their structures were elucidated using MS and NMR experiments such as ^1H, ^{13}C, COSY, HSQC, HMBC, NOESY and ROESY. Satosporin A contained a unique oxo-decalin moiety fused to an eight-membered lactone ring and a disaccharide attached at C16 composed of two β-glucose molecules; satosporin B corresponded to the monosaccharide analogue. Interestingly, these compounds were not produced by any other kitasatosporae studied including a member of the same species, *K. griseola* NRRL B-16229.

The aglycone analogue of the isolated satosporins, though not detected in *K. griseola* fermentations, was desired to assign the absolute stereochemistry at C16. It was obtained via enzymatic deglucosylation by incubation of pure satosporin A with β-glucosidase from almonds and the reaction was monitored by LC-HRMS. The compound was converted to satosporin B within 1 h and fully converted to the aglycone, satosporin C, after 8 d.

The relative configuration of the oxo-decalin moiety of satosporins was determined using NOESY correlations, whereas computational methods were used to compare experimental and

theoretical interatomic distances of the eight-membered lactone ring. The absolute stereochemistry of satosporins was determined using circular dichroism. The modified Mosher's reaction was used to determine the absolute stereochemistry at C16 and the modified Tanaka's derivatization determined that the sugar moieties of satosporin A were both D-glucose. The absolute stereochemistry of satosporins was determined to be 2*R*, 3*R*, 4*S*, 6*R*, 8*R*, 10*S*, 11*R*, 12*S*, 14*R*, 15*R*, 16*R*.

Although satosporin A did not exhibit any significant biological activity against the pathogens and cancer cell lines tested, it was also non-toxic to healthy cells at a concentration of 128 µg/mL. This provides a promising perspective for future applications of this natural product; further investigations regarding their possible biological activities are currently ongoing.

As resistance to current drugs becomes more widely spread and current drugs become ineffective, it is crucial that new chemical entities enter the drug discovery pipeline to be the source of tomorrow's medicines[18]. This research supports previous findings that members of the genus *Kitasatospora* are prolific producers of secondary metabolites and that further investigations of their metabolome may provide additional novel natural products.

7.3 The satosporin biosynthetic gene cluster involves genes encoding multimodular type I polyketide synthases, a cytochrome P450, a glycosyltransferase, a unique fused KS-P450 as well as regulatory and self-resistance genes.

To identify the genetic origin of satosporins in *K. griseola*, data from a degenerate primer PCR-based method was combined with Illumina® and PacBio genome sequencing data. DNA sequences were analysed by antiSMASH, which revealed a 99.7 kb region containing two large PKSs containing one loading and 7 extender modules, consistent with the predicted assembly of the satosporin polyketide backbone. An insertional inactivation of module 1 was generated by homologous recombination and the integration of the apramycin resistance gene was confirmed by PCR and the metabolic profiles from recombinant strains were compared to that from the

wild-type organism. HPLC analysis revealed that the satosporins were no longer present in the crude extracts from recombinant strains, confirming the involvement of the selected genetic region in satosporin biosynthesis.

Sequence analysis, performed using antiSMASH, FramePlot and BLAST®, revealed that the satosporin biosynthetic genes (*ORF1* to *ORF10*) spanned 67 kb and, based on sequence similarities, probably encoded a cytochrome P450 monoxygenase, two large multimodular PKSs, a unique fused ketosynthase-cytochrome P450 monoxygenase, a glycosyltransferase as well as regulatory and self-resistance genes.

A biosynthetic scheme was proposed based on the genetic organization of the *sat* cluster, where the unique tricyclic ring system of satosporins was proposed to originate from a Diels-Alder cyclization following polyketide backbone assembly by the PKSs (SatB and SatC). Two P450 hydroxylations were then proposed to occur (by SatA and/or SatD) to give rise to the aglycone, satosporin C. Finally SatE was proposed to carry out the two glucosylations required to biosynthesize satosporin B and A. Putative pathway specific regulatory genes (*ORF8* and *ORF9*) had high aa sequence homology to other regulatory genes (SARP and unspecific regulatory protein), while *ORF2* had aa sequence homology to efflux proteins and may therefore be involved in regulation and self-resistance to satosporins, respectively.

The inactivation of each biosynthetic gene within the described cluster is essential to confirm the encoding enzymes and the biosynthetic steps they carry out; this is currently being investigated by other members of the laboratory. This research also provided the DNA sequence to allow heterologous production of satosporins. This would confirm that all the necessary genes have been included here, would determine the minimal satosporin biosynthetic gene cluster and could also provide other analogs via combinatorial biosynthesis. As the annotation of secondary metabolite biosynthetic genes relies on nucleotide and amino acid sequence similarities to previously described genes, reporting and characterizing both usual and unusual biosynthetic genes is crucial to developing the most precise gene annotations prediction software. The

characterization of the unique KS-P450 gene (*satD*) will thus be important as it may be the first member of a new enzyme class, whereas characterization of the substrate specificities of the glycosyltransferase may provide a new enzymatic tool for the generation of new non-natural natural products via glycodiversification[170].

7.4 The genome of *K. griseola* MF730-N6 has similar developmental regulatory genes to other kitasatosporae and contains 23 putative secondary metabolite biosynthetic gene clusters.

The genome of *K. griseola* MF730-N6 was sequenced using the Pacific Bioscience RS II platform and, following PCR-based gap closure, resulted in eight contigs covering 7,966,157 bp. The genome was similar to that of other kitasatosporae in its overall G + C content (72.7%), rRNA operons (9 complete + 5 incomplete) and number of protein encoding sequences (7,226).

Kitasatopora griseola has orthologs of almost all developmental regulatory genes (*bld* and *whi* genes) that are usually involved in the complex life cycle of streptomycetes with the exception of *bldB* and *whiJ*. Given the high relatedness of kitasatosporae and streptomycetes, it is not surprising that orthologous genes for important functions are present in members of both genera. In contrast to the *K. setae* genome, orthologs of the ram cluster were identified in *K. griseola*, suggesting that some kitasatosporae may use a protein similar to SapB to facilitate aerial hyphae formation[194].

AntiSMASH analysis revealed 23 putative secondary metabolite biosynthetic gene clusters in the genome of *K. griseola* MF730-N6. The analysis predicted biosynthetic genes leading to the production of seven lantipeptides, six polyketides, three terpenes, two siderophores, two nonribosomal peptide-polyketide hybrids, one nonribosomal peptide, one bacteriocin and one unclassified. Only ten were also present in *K. setae,* suggesting that most natural products produced by kitasatosporae are not distributed throughout members of the genus. As only eight gene clusters could be attributed to known compounds, including

bafilomycin, satosporin and terpentecin, there is the potential to isolate other secondary metabolites from this organism; this could give rise to the isolation of new natural products or it could provide the genetic information for compounds that have already been discovered. Culturing or genetic methods should be employed to induce or increase the production of these numerous cryptic metabolites. The heterologous host developed in Chapter 2 would provide an ideal host as its background metabolite production is minimal and therefore, small amounts of heterologous product may be easily detected by LC-HRMS. The publication of the annotated genome will also allow a more comprehensive comparison to be made between kitasatosporae and streptomycetes.

This research also offers significant advantages for future research involving *K. griseola.* For example, this genome data could quickly provide the biosynthetic origins of two putatively new compounds produced by this organism; performing knock-out experiments would then be necessary to confirm their involvement in the natural product's biosynthesis, saving considerable time and effort.

7.5 Principal component analysis of the metabolomes of twelve kitasatosporae allowed the rapid detection of novel natural products

In order to discover new natural products from "rare" actinomycetes, twelve members of the genus *Kitasatospora* were cultured in two media and extracted with EtOAc. Following LC-HRMS analysis of the crude extracts, the metabolomes were analysed and subjected to principal component analysis according to methods previously developed by members of our lab[24]. The analysis revealed that the metabolomes of *Kitasatospora cystarginea* was unique amongst the kitasatosporae studied. The loadings plot revealed that the ions responsible for this differentiation had *m*/*z* of 357.2022 and 371.2177. Following structure elucidation by Krista Gill (PhD candidate in Kerr laboratory), the compounds highlighted by PCA were confirmed to be new β-lactone containing peptides, cystargolides A and B (manuscript submitted to the Journal

of Natural Products). The same crude extract also resulted in Gill purifying an additional new product, cystargamide, whose structure elucidation was recently published[211]. The PCA also highlighted ions with *m*/*z* 686.4711 to be unique in the crude extract of *K. mediocidica*; AntiBase and preliminary NMR experiments confirmed that this compound was the peptide pepstain A and, thus, was not investigated further.

Although the extracts containing the new compounds (cystargolides and cystargamide) showed no antimicrobial activity against the organisms tested, the crude extracts of *Kitasatospora kifunensis* showed activity against *Candida albicans.* Based on LC-HRMS and NMR data, this activity was probably caused by the known compound naphthomycin A, which was isolated from the crude extract and has reported antifungal activity[213]. This study, however, provided two putatively new compounds from *K. griseola* (compounds with $[M+H]^+$ *m*/*z* 542.3229 and 556.3389) with possible biological activity against MRSA and *S. warneri.* Gill is currently working towards purifying sufficient amounts to elucidate their structure.

This research supports the use of LC-HRMS based metabolic profiling and PCA to rapidly identify new compounds and confirms the use of studying "rare" actinomycetes such as kitasatosporae for the discovery of novel natural products. Future studies involving "rare" actinomycetes for the discovery of new natural products could include a more comprehensive collection of kitasatosporae from both culture collections and from unique environments as strain-specific natural products can evolve in members of the same species, evidenced here by terpentecin and satosporin production in *K. griseola* MF730-N6. Genera to incorporate in future analyses could include *Salinispora*, *Actinosynnema*, *Kibdelosporangium*, *Microellobosporia*, *Chainia*, etc.[217] and the use of various fermentation conditions could be used to trigger the production of as many natural products as possible.

7.6 Concluding remarks

This thesis work has provided advancements in various disciplines within natural product research in the form of new/accessible molecular tools, new natural products with unique chemical structures and genetic and metabolomic information from kitasatosporae. Other studies have already stemed from research within this thesis and, hopefully, natural products drug discovery from kitasatosporae will continue beyond the work that I have accomplished.

REFERENCES

1. Diminic, J., A. Starcevic, M. Lisfi, D. Baranasic, R. Gacesa, D. Hranueli, P.F. Long, J. Cullum, and J. Zucko, Evolutionary concepts in natural products discovery: what actinomycetes have taught us. *J. Ind. Microbiol. Biotechnol.*, 2014. **41**(2): 211-217.

2. Stone, M.J. and D.H. Williams, On the evolution of functional secondary metabolites (natural products). *Mol. Microbiol.*, 1992. **6**(1): 29-34.

3. Paduch, R., M. Kandefer-Szerszen, M. Trytek, and J. Fiedurek, Terpenes: substances useful in human healthcare. *Arch. Immunol. Ther. Exp. (Warsz).* 2007. **55**(5): 315-327.

4. Ajikumar, P.K., K. Tyo, S. Carlsen, O. Mucha, T.H. Phon, and G. Stephanopoulos, Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharm.*, 2008. **5**(2): 167-190.

5. Dairi, T., Studies on biosynthetic genes and enzymes of isoprenoids produced by actinomycetes. *J. Antibiot.*, 2005. **58**(4): 227-43.

6. Kirby, J. and J.D. Keasling, Metabolic engineering of microorganisms for isoprenoid production. *Nat. Prod. Rep.*, 2008. **25**(4): 656-661.

7. Hranueli, D., N. Peric, B. Borocicka, S. Bogdan, J. Cullum, P. Waterman, and I. Hunter, Molecular biology of polyketide biosynthesis. *Food Technol. Biotechnol.*, 2001. **39**(3): 203-213.

8. Khosla, C., Structures and mechanisms of polyketide synthases. *J. Org. Chem.*, 2009. **74**(17): 6416-6420.

9. Hertweck, C., The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed. Engl.*, 2009. **48**(26): 4688-716.

10. Fischbach, M.A. and C.T. Walsh, Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.*, 2006. **106**(8): 3468-96.

11. Dewick, P.M., Medicinal natural products: a biosynthetic approach. Chichester, England: Wiley; 2002. 550 p.

12. Arnison, P.G., M.J. Bibb, G. Bierbaum, A.A. Bowers, T.S. Bugni, G. Bulaj, J.A. Camarero, D.J. Campopiano, G.L. Challis, J. Clardy, P.D. Cotter, D.J. Craik, M. Dawson, E. Dittmann, S. Donadio, P.C. Dorrestein, K.-D. Entian, M.A. Fischbach, J.S. Garavelli, U. Goransson, C.W. Gruber, D.H. Haft, T.K. Hemscheidt, C. Hertweck, C. Hill, A.R. Horswill, M. Jaspars, W.L. Kelly, J.P. Klinman, O.P. Kuipers, A.J. Link, W. Liu, M.A. Marahiel, D.A. Mitchell, G.N. Moll, B.S. Moore, R. Muller, S.K. Nair, I.F. Nes, G.E. Norris, B.M. Olivera, H. Onaka, M.L. Patchett, J. Piel, M.J.T. Reaney, S. Rebuffat, R.P. Ross, H.-G. Sahl, E.W. Schmidt, M.E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein, R.D. Sussmuth, J.R. Tagg, G.-L. Tang, A.W. Truman, J.C. Vederas, C.T. Walsh, J.D. Walton, S.C. Wenzel, J.M. Willey, and W.A. van der Donk, Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.*, 2013. **30**(1): 108-160.

13. Gershenzon, J. and N. Dudareva, The function of terpene natural products in the natural world. *Nat. Chem. Biol.*, 2007. **3**(7): 408-414.

14. Rath, C.M., B. Janto, J. Earl, A. Ahmed, F.Z. Hu, L. Hiller, M. Dahlgren, R. Kreft, F. Yu, J.J. Wolff, H.K. Kweon, M.A. Christiansen, K. Hakansson, R.M. Williams, G.D. Ehrlich, and D.H.

Sherman, Meta-omic characterization of the marine invertebrate microbial consortium that produces the chemotherapeutic natural product ET-743. *ACS Chem. Biol.*, 2011. **6**(11): 1244-56.

15. Piel, J., Metabolites from symbiotic bacteria. *Nat. Prod. Rep.*, 2009. **26**(3): 338-362.

16. Piel, J., D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani, and S. Matsunaga, Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc. Natl. Acad. Sci. U. S. A.*, 2004. **101**(46): 16222-16227.

17. Gerwick, W.H. and B.S. Moore, Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chem. Biol.*, 2012. **19**(1): 85-98.

18. Cragg, G.M. and D.J. Newman, Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta*, 2013. **1830**(6): 3670-95.

19. Koehn, F.E. and G.T. Carter, The evolving role of natural products in drug discovery. *Nat. Rev. Drug Discov.*, 2005. **4**(3): 206-20.

20. Newman, D.J. and G.M. Cragg, Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.*, 2012. **75**(3): 311-335.

21. Paul, S.M., D.S. Mytelka, C.T. Dunwiddie, C.C. Persinger, B.H. Munos, S.R. Lindborg, and A.L. Schacht, How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.*, 2010. **9**(3): 203-214.

22. Koehn, F.E. and G.T. Carter, The evolving role of natural products in drug discovery. *Nat. Rev. Drug Discov.*, 2005. **4**(3): 206-20.

23. Lang, G., N.A. Mayhudin, M.I. Mitova, L. Sun, S. van der Sar, J.W. Blunt, A.L. Cole, G. Ellis, H. Laatsch, and M.H. Munro, Evolving trends in the dereplication of natural product extracts: new methodology for rapid, small-scale investigation of natural product extracts. *J. Nat. Prod.*, 2008. **71**(9): 1595-9.

24. Forner, D., F. Berrué, H. Correa, K. Duncan, and R.G. Kerr, Chemical dereplication of marine actinomycetes by liquid chromatography-high resolution mass spectrometry profiling and statistical analysis. Submitted. *Anal. Chim. Acta*, 2013. **805**: 70-79.

25. Watrous, J., P. Roach, T. Alexandrov, B.S. Heath, J.Y. Yang, R.D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J.M. Raaijmakers, B.S. Moore, J. Laskin, N. Bandeira, and P.C. Dorrestein, Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. U. S. A.*, 2012. **109**(26): E1743-52.

26. Bull, A.T. and J.E.M. Stach, Marine actinobacteria: new opportunities for natural product search and discovery. *Trends Microbiol.*, 2007. **15**(11): 491-499.

27. De Vries, D.J. and M.R. Hall, Marine biodiversity as a source of chemical diversity. *Drug Dev. Res.*, 1994. **33**(2): 161-173.

28. Genilloud, O., I. Gonzalez, O. Salazar, J. Martin, J. Ruben Tormo, and F. Vicente, Current approaches to exploit actinomycetes as a source of novel natural products. *J. Ind. Microbiol. Biotechnol.*, 2011. **38**(3): 375-389.

29. Rebets, Y., E. Brotz, B. Tokovenko, and A. Luzhetskyy, Actinomycetes biosynthetic potential: how to bridge *in silico* and *in vivo*? *J. Ind. Microbiol. Biotechnol.*, 2014. **41**(2): 387-402.

30. Bentley, S.D., K.F. Chater, A.M. Cerdeno-Tarraga, G.L. Challis, N.R. Thomson, K.D. James, D.E. Harris, M.A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C.W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M.A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B.G. Barrell, J. Parkhill, and D.A. Hopwood, Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, 2002. **417**(6885): 141-147.

31. Craney, A., S. Ahmed, and J. Nodwell, Towards a new science of secondary metabolism. *J. Antibiot.*, 2013. **66**(7, Sp. Iss. SI): 387-400.

32. Lautru, S., R.J. Deeth, L.M. Bailey, and G.L. Challis, Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nat. Chem. Biol.*, 2005. **1**(5): 265-269.

33. Bibb, M.J., Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.*, 2005. **8**(2): 208-215.

34. Aroonsri, A., S. Kitani, H. Ikeda, and T. Nihira, Kitasetaline, a novel beta-carboline alkaloid from *Kitasatospora setae* NBRC 14216(T). *J. Biosci. Bioeng.*, 2012. **114**(1): 56-58.

35. Quail, M.A., M. Smith, P. Coupland, T.D. Otto, S.R. Harris, T.R. Connor, A. Bertoni, H.P. Swerdlow, and Y. Gu, A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 2012. **13**: 341-354.

36. Mardis, E.R., Next-generation sequencing platforms. *Annu. Rev. Anal. Chem.*, 2013. **6**: 287-303.

37. Schneider, G.F. and C. Dekker, DNA sequencing with nanopores. *Nat Biotech*, 2012. **30**(4): 326-328.

38. Blin, K., M.H. Medema, D. Kazempour, M.A. Fischbach, R. Breitling, E. Takano, and T. Weber, antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res*, 2013. **41**(W1): W204-W212.

39. Fortman, J.L. and D.H. Sherman, Utilizing the power of microbial genetics to bridge the gap between the promise and the application of marine natural products. *Chembiochem*, 2005. **6**(6): 960-78.

40. Look, S.A., W. Fenical, R.S. Jacobs, and J. Clardy, The pseudopterosins: anti-inflammatory and analgesic natural products from the sea whip *Pseudopterogorgia elisabethae*. *Proc. Natl. Acad. Sci. U. S. A.*, 1986. **83**(17): 6238-40.

41. Lindel, T., P.R. Jensen, W. Fenical, B.H. Long, A.M. Casazza, J. Carboni, and C.R. Fairchild, Eleutherobin, a new cytotoxin that mimics paclitaxel (Taxol) by stabilizing microtubules. *J. Am. Chem. Soc.*, 1997. **119**(37): 8744-8745.

42. Kerr, R.G., A.C. Kohl, and T.A. Ferns, Elucidation of the biosynthetic origin of the anti-inflammatory pseudopterosins. *J. Ind. Microbiol. Biotechnol.*, 2006. **33**: 532-538.

43. Kijjoa, A. and P. Sawangwong, Drugs and cosmetics from the sea. *Mar. Drugs*, 2004. **2**(2): 73-82.

44. Long, B.H., J.M. Carboni, A.J. Wasserman, L.A. Cornell, A.M. Casazza, P.R. Jensen, T. Lindel, W. Fenical, and C.R. Fairchild, Eleutherobin, a novel cytotoxic agent that induces tubulin polymerization, is similar to paclitaxel (Taxol). *Cancer Res.*, 1998. **58**(6): 1111-5.

45. Berrué, F., M.W.B. McCulloch, and R.G. Kerr, Marine diterpene glycosides. *Bioorg. Med. Chem.*, 2011. **19**(22): 6702-6719.

46. Lee, S.Y., H.U. Kim, J.H. Park, J.M. Park, and T.Y. Kim, Metabolic engineering of microorganisms: general strategies and drug production. *Drug Discov. Today*, 2009. **14**(1-2): 78-88.

47. Olano, C., F. Lombo, C. Mendez, and J.A. Salas, Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. *Metab. Eng.*, 2008. **10**(5): 281-292.

48. Zhang, H., Y. Wang, and B.A. Pfeifer, Bacterial hosts for natural product production. *Mol. Pharm.*, 2008. **5**(2): 212-225.

49. Galm, U. and B. Shen, Expression of biosynthetic gene clusters in heterologous hosts for natural product production and combinatorial biosynthesis. *Expert Opin. Drug. Discov.*, 2006. **1**(5): 409-437.

50. Binz, T.M., S.C. Wenzel, H.J. Schnell, A. Bechthold, and R. Muller, Heterologous expression and genetic engineering of the phenalinolactone biosynthetic gene cluster by using Red/ET recombineering. *Chembiochem*, 2008. **9**(3): 447-454.

51. Ichinose, K., M. Ozawa, K. Itou, K. Kunieda, and Y. Ebizuka, Cloning, sequencing and heterologous expression of the medermycin biosynthetic gene cluster of *Streptomyces* sp. AM-7161: towards comparative analysis of the benzoisochromanequinone gene clusters. *Microbiology*, 2003. **149**: 1633-1645.

52. Hong, S.T., J.R. Carney, and S.J. Gould, Cloning and heterologous expression of the entire gene clusters for PD 116740 from *Streptomyces* strain WP 4669 and tetrangulol and tetrangomycin from *Streptomyces rimosus* NRRL 3016. *J. Bacteriol.*, 1997. **179**(2): 470-476.

53. Kim, S.Y., P. Zhao, M. Igarashi, R. Sawa, T. Tomita, M. Nishiyama, and T. Kuzuyama, Cloning and heterologous expression of the cyclooctatin biosynthetic gene cluster afford a diterpene cyclase and two P450 hydroxylases. *Chem. Biol.*, 2009. **16**(7): 736-743.

54. Julia, P., L. Xiang, A. Whiting, L. Mohammed, T. Gibson, C.J. Silva, P. Brian, J. Davies, V. Miao, S.K. Wrigley, and R.H. Baltz, Heterologous production of daptomycin in *Streptomyces lividans*. *J. Ind. Microbiol. Biotechnol.*, 2006. **33**(2): 121-128.

55. Smanski, M.J., J. Casper, R.M. Peterson, Z. Yu, S.R. Rajski, and B. Shen, Expression of the platencin biosynthetic gene cluster in heterologous hosts yielding new platencin congeners. *J. Nat. Prod.*, 2012. **75**(12): 2158-2167.

56. Martin, V.J.J., D.J. Pitera, S.T. Withers, J.D. Newman, and J.D. Keasling, Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.*, 2003. **21**(7): 796-802.

57. Leonard, E., P.K. Ajikumar, K. Thayer, W.-H. Xiao, J.D. Mo, B. Tidor, G. Stephanopoulos, and K.L.J. Prather, Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc. Natl. Acad. Sci. U. S. A.*, 2010. **107**(31): 13654-13659.

58. Paddon, C.J., P.J. Westfall, D.J. Pitera, K. Benjamin, K. Fisher, D. McPhee, M.D. Leavell, A. Tai, A. Main, D. Eng, D.R. Polichuk, K.H. Teoh, D.W. Reed, T. Treynor, J. Lenihan, M. Fleck, S. Bajad, G. Dang, D. Dengrove, D. Diola, G. Dorin, K.W. Ellens, S. Fickes, J. Galazzo, S.P. Gaucher, T. Geistlinger, R. Henry, M. Hepp, T. Horning, T. Iqbal, H. Jiang, L. Kizer, B. Lieu, D.

Melis, N. Moss, R. Regentin, S. Secrest, H. Tsuruta, R. Vazquez, L.F. Westblade, L. Xu, M. Yu, Y. Zhang, L. Zhao, J. Lievense, P.S. Covello, J.D. Keasling, K.K. Reiling, N.S. Renninger, and J.D. Newman, High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 2013. **496**(7446): 528-32.

59. Tsuruta, H., C.J. Paddon, D. Eng, J.R. Lenihan, T. Horning, L.C. Anthony, R. Regentin, J.D. Keasling, N.S. Renninger, and J.D. Newman, High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in *Escherichia coli*. *PLoS ONE*, 2009. **4**(2): e4489.

60. Paddon, C.J. and J.D. Keasling, Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.*, 2014. **12**(5): 355-67.

61. Demain, A. and J. Adrio, Strain improvement for production of pharmaceuticals and other microbial metabolites by fermentation, in Natural compounds as drugs F. Petersen and R. Amstutz, Editors. Basel, Switzerland: Birkhäuser 2008. p. 251-289.

62. Demain, A.L., From natural products discovery to commercialization: a success story. *J. Ind. Microbiol. Biotechnol.*, 2006. **33**(7): 486-95.

63. Zhang, H., B.A. Boghigian, J. Armando, and B.A. Pfeifer, Methods and options for the heterologous production of complex natural products. *Nat. Prod. Rep.*, 2011. **28**(1): 125-51.

64. Dairi, T., Y. Hamano, T. Kuzuyama, N. Itoh, K. Furihata, and H. Seto, Eubacterial diterpene cyclase genes essential for production of the isoprenoid antibiotic terpentecin. *J. Bacteriol.*, 2001. **183**(20): 6085-6094.

65. Tetzlaff, C.N., Z. You, D.E. Cane, S. Takamatsu, S. Omura, and H. Ikeda, A gene cluster for biosynthesis of the sesquiterpenoid antibiotic pentalenolactone in *Streptomyces avermitilis*. *Biochemistry*, 2006. **45**(19): 6179-6186.

66. Maury, J., M.A. Asadollahi, K. Moller, A. Clark, and J. Nielsen, Microbial isoprenoid production: an example of green chemistry through metabolic engineering, in Biotechnology for the future. Heidelberg, Germany: Springer-Verlag; 2005. p. 19-51.

67. Tamamura, T., T. Sawa, K. Isshiki, T. Masuda, Y. Homma, H. Iinuma, H. Naganawa, M. Hamada, T. Takeuchi, and H. Umezawa, Isolation and characterization of terpentecin, a new antitumor antibiotic. *J. Antibiot.*, 1985. **38**(12): 1664-1669.

68. Sambrook, J. and D.W. Russell, Molecular cloning-a laboratory manual. 3rd ed. New York: Cold Spring Harbor Lab Press; 2001. 2100 p.

69. Bierman, M., R. Logan, K. Obrien, E.T. Seno, R.N. Rao, and B.E. Schoner, Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene*, 1992. **116**(1): 43-49.

70. Flett, F., V. Mersinias, and C.P. Smith, High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett.*, 1997. **155**(2): 223-229.

71. Kieser, T., M.J. Bibb, M.J. Buttner, K.F. Chater, and D.A. Hopwood, Practical *Streptomyces* genetics. Norwich, England: John Innes Foundation; 2000. 613 p.

72. Okamoto, S., T. Taguchi, K. Ochi, and K. Ichinose, Biosynthesis of actinorhodin and related antibiotics: discovery of alternative routes for quinone formation encoded in the act gene cluster. *Chem. Biol.*, 2009. **16**(2): 226-236.

73. Cerdeno, A.M., M.J. Bibb, and G.L. Challis, Analysis of the prodiginine biosynthesis gene cluster of *Streptomyces coelicolor* A3(2): new mechanisms for chain initiation and termination in modular multienzymes. *Chem. Biol.*, 2001. **8**(8): 817-829.

74. Borodina, I., J. Siebring, J. Zhang, C.P. Smith, G. van Keulen, L. Dijkhuizen, and J. Nielsen, Antibiotic overproduction in *Streptomyces coelicolor* A3(2) mediated by phosphofructokinase deletion. *J. Biol. Chem.*, 2008. **283**(37): 25186-25199.

75. Hamano, Y., T. Dairi, M. Yamamoto, T. Kawasaki, K. Kaneda, T. Kuzuyama, N. Itoh, and H. Seto, Cloning of a gene cluster encoding enzymes responsible for the mevalonate pathway from a terpenoid-antibiotic-producing *Streptomyces* strain. *Biosci. Biotechnol. Biochem.*, 2001. **65**(7): 1627-1635.

76. Wilkinson, C.J., Z.A. Hughes-Thomas, C.J. Martin, I. Bohm, T. Mironenko, M. Deacon, M. Wheatcroft, G. Wirtz, J. Staunton, and P.F. Leadlay, Increasing the efficiency of heterologous promoters in actinomycetes. *J. Mol. Microbiol. Biotechnol.*, 2002. **4**(4): 417-426.

77. Zhang, W., J.L. Fortman, J.C. Carlson, J. Yan, Y. Liu, F. Bai, W. Guan, J. Jia, T. Matainaho, D.H. Sherman, and S. Li, Characterization of the bafilomycin biosynthetic gene cluster from *Streptomyces lohii*. *ChemBioChem*, 2013. **14**(3): 301-306.

78. Choi, S.-U., C.-K. Lee, Y.-I. Hwang, H. Kinoshita, and T. Nihira, Intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to *Kitasatospora setae*, a bafilomycin B1 producer. *Arch. Microbiol.*, 2004. **181**(4): 294-8.

79. Wright, L.F. and D.A. Hopwood, Actinorhodin is a chromosomally-determined antibiotic in *Streptomyces coelicolor* A3(2). *J. Gen. Microbiol.*, 1976. **96**(OCT): 289-297.

80. Rudd, B.A.M. and D.A. Hopwood, A pigmented mycelial antibiotic in *Streptomyces coelicolor* - control by a chromosomal gene-cluster. *J. Gen. Microbiol.*, 1980. **119**(AUG): 333-340.

81. Zar, J.H., Biostatistical analysis. Fourth ed. Upper saddle river, New Jersey: Prentice Hall; 1999. 663 p.

82. Martinez, A., S.J. Kolvek, C.L.T. Yip, J. Hopke, K.A. Brown, I.A. MacNeil, and M.S. Osburne, Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl. Environ. Microbiol.*, 2004. **70**(4): 2452-63.

83. MacNeil, D.J., K.M. Gewain, C.L. Ruby, G. Dezeny, P.H. Gibbons, and T. MacNeil, Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene*, 1992. **111**(1): 61-68.

84. Davis, J.R. and J.K. Sello, Regulation of genes in *Streptomyces* bacteria required for catabolism of lignin-derived aromatic compounds. *Appl. Microbiol. Biotechnol.*, 2010. **86**(3): 921-929.

85. Buttner, M.J., K.F. Chater, and M.J. Bibb, Cloning, disruption, and transcriptional analysis of three RNA polymerase sigma factor genes of *Streptomyces coelicolor* A3(2). *J. Bacteriol.*, 1990. **172**(6): 3367-78.

86. Kämpfer, P., Genus *incertae sedis* I. *Kitasatospora*, in Bergey's manual of systematic bacteriology, Volume 5: The Actinobacteria. New York: Springer; 2012. p. 1768-1777.

87. Andersen, N.R. and P.R. Rasmussen, The constitution of clerocidin a new antibiotic isolated from *Oidiodendron truncatum*. *Tetrahedron Lett.*, 1984. **25**(4): 465-468.

88. Hamano, U., T. Takeuchi, M. Hamada, H. Naganawa, Y. takahashi, H. Iinuma, T. Sawa, T. Tamamura, and K. Isshiki, A novel anti-tumor and antimicrobial compound, its microbiological preparation and its use as medicament. *European Patent Office*, 1986. **EP0205981A2**.

89. Nakano, H., S. Kawada, Y. Uosaki, Y. Saito, K. Gomi, and T. Iwazaki, Process for producing diterpene compounds. *United states patent office*, 1992. **5,151,352**.

90. Omura, S., K. Otoguro, T. Nishikiori, R. Oiwa, and Y. Iwai, Setamycin a new antibiotic. *J. Antibiot.*, 1981. **34**(10): 1253-1256.

91. Kittendorf, J.D. and D.H. Sherman, The methymycin/pikromycin pathway: a model for metabolic diversity in natural product biosynthesis. *Bioorg. Med. Chem.*, 2009. **17**(6): 2137-46.

92. Berdy, J., Bioactive microbial metabolites - a personal view. *J. Antibiot.*, 2005. **58**(1): 1-26.

93. Demain, A.L. and S. Sanchez, Microbial drug discovery: 80 years of progress. *J. Antibiot.*, 2009. **62**(1): 5-16.

94. Omura, S., Y. Takahashi, Y. Iwai, and H. Tanaka, *Kitasatosporia* new-genus of the order Actinomycetales. *J. Antibiot.*, 1982. **35**(8): 1013-1019.

95. Wellington, E.M.H., E. Stackebrandt, D. Sanders, J. Wolstrup, and N.O.G. Jorgensen, Taxonomic status of *Kitasatosporia* and proposed unification with *Streptomyces* on the basis of phenotypic and 16s rRNA analysis and emendation of *Streptomyces* Waksman and Henrici 1943, 339. *Int. J. Syst. Bacteriol.*, 1992. **42**(1): 156-160.

96. Zhang, Z., Y. Wang, and J. Ruan, A proposal to revive the genus *Kitasatospora* (Omura, Takahashi, Iwai, and Tanaka 1982). *Int. J. Syst. Bacteriol.*, 1997. **47**(4): 1048-1054.

97. Girard, G., B.A. Traag, V. Sangal, N. Mascini, P.A. Hoskisson, M. Goodfellow, and G.P. van Wezel, A novel taxonomic marker that discriminates between morphologically complex actinomycetes. *Open Biol.*, 2013. **3**(10): 130073-130073.

98. Takahashi, Y. and S. Omura, Isolation of new actinomycete strains for the screening of new bioactive compounds. *J. Gen. Appl. Microbiol.*, 2003. **49**(3): 141-154.

99. Maekawa, K., K. Toume, and M. Ishibashi, Isolation of new fuzanins, carbamate-containing natural products, from *Kitasatospora* sp IFM10917. *J. Antibiot.*, 2010. **63**(7): 385-388.

100. Heine, D., K. Martin, and C. Hertweck, Genomics-guided discovery of endophenazines from *Kitasatospora* sp. HKI 714. *J. Nat. Prod.*, 2014. **77**(4): 1083-1087.

101. Shafiee, A., A. Tsipouras, A. Bouffard, J.C. Onishi, Z. Guan, and H. Motamedi, Enzymatic deglycosylation of enfumafungin, a triterpene glycoside natural product, and its chemically synthesized analogues. *J. Mol. Catal. B Enzym.*, 2001. **16**(1): 27-32.

102. Gottlieb, H.E., V. Kotlyar, and A. Nudelman, NMR chemical shifts of common laboratory solvents as trace impurities. *J. Org. Chem.*, 1997. **62**(21): 7512-7515.

103. Stewart, J.J.P., Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.*, 1989. **10**(2): 209-220.

104. Lee, C., W. Yang, and R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 1988. **37**(2): 785-789.

105. Becke, A.D., Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 1993. **98**(7): 5648-5652.

106. Tomasi, J., B. Mennucci, and R. Cammi, Quantum mechanical continuum solvation models. *Chem. Rev.*, 2005. **105**(8): 2999-3093.

107. Gross, E.K.U., J.F. Dobson, and M. Petersillka, Density functional theory of time-dependent phenomena, in Density functional theory, R.F. Nalewajski, Editor. Berlin: Springer; 1996. p. 81-172.

108. Latypov, S.K., J.M. Seco, E. Quinoa, and R. Riguera, Are both the *(R)*- and the *(S)*-MPA esters really needed for the assignment of the absolute configuration of secondary alcohols by NMR? The use of a single derivative. *J. Am. Chem. Soc.*, 1998. **120**: 877-882.

109. Cairns, J.R.K. and A. Esen, beta-Glucosidases. *Cell. Mol. Life Sci.*, 2010. **67**(20): 3389-3405.

110. Hwang, J.Y., H.S. Kim, S.H. Kim, H.R. Oh, and D.H. Nam, Organization and characterization of a biosynthetic gene cluster for bafilomycin from *Streptomyces griseus* DSM 2608. *AMB Express*, 2013. **3**(1): 24-24.

111. Quiros, L.M., R.J. Carbajo, A.F. Brana, and J.A. Salas, Glycosylation of macrolide antibiotics. Purification and kinetic studies of a macrolide glycosyltransferase from *Streptomyces antibioticus*. *J. Biol. Chem.*, 2000. **275**(16): 11713-20.

112. Bode, H.B. and R. Muller, The impact of bacterial genomics on natural product research. *Angew. Chem.*, 2005. **44**(42): 6828-6846.

113. Fang, X.-P., J.E. Anderson, X.-X. Qiu, J.F. Kozlowski, C.-J. Chang, and J.L. McLaughlin, Gonioheptolides A and B: novel eight-membered-ring lactones from *Goniothalamus giganteus* (Annonaceae). *Tetrahedron*, 1993. **49**(8): 1563-1570.

114. Singh, S., G.N. Phillips, Jr., and J.S. Thorson, The structural biology of enzymes involved in natural product glycosylation. *Nat. Prod. Rep.*, 2012. **29**(10): 1201-37.

115. Losey, H.C., M.W. Peczuh, Z. Chen, U.S. Eggert, S.D. Dong, I. Pelczer, D. Kahne, and C.T. Walsh, Tandem action of glycosyltransferases in the maturation of vancomycin and teicoplanin aglycones: novel glycopeptides. *Biochemistry*, 2001. **40**(15): 4745-55.

116. Onaka, H., S. Taniguchi, Y. Igarashi, and T. Furumai, Characterization of the biosynthetic gene cluster of rebeccamycin from *Lechevalieria aerocolonigenes* ATCC 39243. *Biosci. Biotechnol. Biochem.*, 2003. **67**(1): 127-38.

117. Flatt, P.M., X. Wu, S. Perry, and T. Mahmud, Genetic insights into pyralomicin biosynthesis in *Nonomuraea spiralis* IMC A-0156. *J. Nat. Prod.*, 2013. **76**(5): 939-46.

118. Pavia, D.L., G.M. Lampman, and G.S. Kriz, Introduction to spectroscopy-a guide for students of organic chemistry. 3rd ed. Unites States of America: Thomson Learning; 2001. 579 p.

119. Kwan, D.H. and F. Schulz, The stereochemistry of complex polyketide biosynthesis by modular polyketide synthases. *Molecules*, 2011. **16**(7): 6092-115.

120. Bringmann, G., T. Bruhn, K. Maksimenka, and Y. Hemberger, The assignment of absolute stereostructures through quantum chemical circular dichroism calculations. *European J. Org. Chem.*, 2009. **2009**(17): 2717-2727.

121. Dai, J., K. Krohn, B. Elsässer, U. Flörke, S. Draeger, B. Schulz, G. Pescitelli, P. Salvadori, S. Antus, and T. Kurtán, Metabolic products of the endophytic fungus *Microsphaeropsis* sp. from *Larix decidua*. *European J. Org. Chem.*, 2007. **2007**(29): 4845-4854.

122. Wang, Y., G. Raabe, C. Repges, and J. Fleischhauer, Time-dependent density functional theory calculations on the chiroptical properties of rubroflavin: determination of its absolute configuration by comparison of measured and calculated CD spectra. *Int. J. Quantum Chem.*, 2003. **93**(4): 265-270.

123. Neiss, C., P. Saalfrank, M. Parac, and S. Grimme, Quantum chemical calculation of excited states of flavin-related molecules. *J. Phys. Chem. A.*, 2002. **107**(1): 140-147.

124. Diedrich, C. and S. Grimme, Systematic investigation of modern quantum chemical methods to predict electronic circular dichroism spectra. *J. Phys. Chem. A.*, 2003. **107**(14): 2524-2539.

125. Bringmann, G. and S. Busemann, The quantum chemical calculation of CD spectra: the absolute configuration of chiral compounds from natural or synthetic origin in Natural product analysis: chromatography, spectroscopy, biological testing, P. Schreier, M. Herderich, H.-U. Humpf, and W. Schwab, Editors. Wiesbaden, Germany: Springer Vieweg; 1998. p. 195-211.

126. Kawada, S., Y. Yamashita, K. Ochiai, K. Ando, T. Iwasaki, T. Takiguchi, and H. Nakano, Terpentecin and uct4b, new family of topoisomerase-II targeting antitumor antibiotics produced by *Streptomyces* - producing organism, fermentation and large-scale purification. *J. Antibiot.*, 1995. **48**(3): 211-216.

127. Groth, I., C. Rodriguez, B. Schuetze, P. Schmitz, E. Leistner, and M. Goodfellow, Five novel *Kitasatospora* species from soil: *Kitasatospora arboriphila* sp. nov., *K. gansuensis* sp. nov., *K. nipponensis* sp. nov., *K. paranensis* sp. nov. and *K. terrestris* sp. nov. *Int. J. Syst. Evol. Microbiol.*, 2004. **54**(Part 6): 2121-2129.

128. Rude, M.A. and C. Khosla, Engineered biosynthesis of polyketides in heterologous hosts. *Chem. Eng. Sci.*, 2004. **59**(22-23): 4693-4701.

129. Arens, J.C., F. Berrue, J.K. Pearson, and R.G. Kerr, Isolation and structure elucidation of satosporin A and B: new polyketides from *Kitasatospora griseola*. *Org. Lett.*, 2013. **15**(15): 3864-7.

130. Vieira, J. and J. Messing, The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene*, 1982. **19**(3): 259-68.

131. Hall, T.A., BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.*, 1999. **41**: 95-98.

132. Ishikawa, J. and K. Hotta, FramePlot: A new implementation of the Frame analysis for predicting protein-coding regions in bacterial DNA with a high G+C content. *FEMS Microbiol. Lett.*, 1999. **174**(2): 251-253.

133. Ayuso-Sacido, A. and O. Genilloud, New PCR primers for the screening of NRPS and PKS-I systems in actinomycetes: detection and distribution of these biosynthetic gene sequences in major taxonomic groups. *Microb. Ecol.*, 2005. **49**(1): 10-24.

134. Anand, S., M.V. Prasad, G. Yadav, N. Kumar, J. Shehara, M.Z. Ansari, and D. Mohanty, SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Res*, 2010. **38**(Web Server issue): W487-96.

135. Clark, J.M., Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res*, 1988. **16**(20): 9677-86.

136. Ichikawa, N., A. Oguchi, H. Ikeda, J. Ishikawa, S. Kitani, Y. Watanabe, S. Nakamura, Y. Katano, E. Kishi, M. Sasagawa, A. Ankai, S. Fukui, Y. Hashimoto, S. Kamata, M. Otoguro, S. Tanikawa, T. Nihira, S. Horinouchi, Y. Ohnishi, M. Hayakawa, T. Kuzuyama, A. Arisawa, F. Nomoto, H. Miura, Y. Takahashi, and N. Fujita, Genome sequence of *Kitasatospora setae* NBRC 14216T: an evolutionary snapshot of the family Streptomycetaceae. *DNA Res.*, 2010. **17**(6): 393-406.

137. Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, Basic local alignment search tool. *J. Mol. Biol.*, 1990. **215**(3): 403-10.

138. Ferrarini, M., M. Moretto, J.A. Ward, N. Urbanovski, V. Stevanovi, L. Giongo, R. Viola, D. Cavalieri, R. Velasco, A. Cestaro, and D.J. Sargent, An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. *BMC Genomics*, 2013. **14**(1): 670.

139. Djambazian, H., L. Létourneau, P. Bérubé, G. Geneau, P. Marquis, G. Leveque, A. Staffa, A. Montpetit, R. Sladek, and K. Dewar, Towards finished microbial genome assemblies using Pacific Biosciences. 2012, McGill University and Genome Quebec Innovation Centre (Technical report). 4.

140. Osbourn, A., Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet.*, 2010. **26**(10): 449.

141. Rost, B., Twilight zone of protein sequence alignments. *Protein Eng.*, 1999. **12**(2): 85-94.

142. Long, P.F., C.J. Wilkinson, C.P. Bisang, J. Cortes, N. Dunster, M. Oliynyk, E. McCormick, H. McArthur, C. Mendez, J.A. Salas, J. Staunton, and P.F. Leadlay, Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase. *Mol. Microbiol.*, 2002. **43**(5): 1215-25.

143. Yadav, G., R.S. Gokhale, and D. Mohanty, Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.*, 2003. **328**(2): 335-63.

144. Young, J. and R.E. Taylor, Evolution of polyketides: post-PKS processing in the formation of spiroketals. *Nat. Prod. Rep.*, 2008. **25**(4): 651-655.

145. Keatinge-Clay, A.T., A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.*, 2007. **14**(8): 898-908.

146. Keatinge-Clay, A.T., The structures of type I polyketide synthases. *Nat. Prod. Rep.*, 2012. **29**(10): 1050-1073.

147. Broadhurst, R.W., D. Nietlispach, M.P. Wheatcroft, P.F. Leadlay, and K.J. Weissman, The structure of docking domains in modular polyketide synthases. *Chem. Biol.*, 2003. **10**(8): 723-31.

148. Stocking, E.M. and R.M. Williams, Chemistry and biology of biosynthetic Diels-Alder reactions. *Angew. Chem. Int. Ed. Engl.*, 2003. **42**(27): 3078-115.

149. Kasahara, K., T. Miyamoto, T. Fujimoto, H. Oguri, T. Tokiwano, H. Oikawa, Y. Ebizuka, and I. Fujii, Solanapyrone synthase, a possible Diels-Alderase and iterative type I polyketide synthase encoded in a biosynthetic gene cluster from *Alternaria solani*. *Chembiochem*, 2010. **11**(9): 1245-52.

150. Cane, D.E., W. Tan, and W.R. Ott, Nargenicin biosynthesis. Incorporation of polyketide chain elongation intermediates and support for a proposed intramolecular Diels-Alder cyclization. *J. Am. Chem. Soc.*, 1993. **115**(2): 527-535.

151. Oikawa, H., A. Ichihara, and S. Sakamura, Biosynthetic study of betaenone B: origin of the oxygen atoms and accumulation of a deoxygenated intermediate using P-450 inhibitor. *J. Chem. Soc., Chem. Commun.*, 1988(9): 600-602.

152. Auclair, K., A. Sutherland, J. Kennedy, D.J. Witter, J.P. Van den Heever, C.R. Hutchinson, and J.C. Vederas, Lovastatin nonaketide synthase catalyzes an intramolecular Diels−Alder reaction of a substrate analogue. *J. Am. Chem. Soc.*, 2000. **122**(46): 11519-11520.

153. Ose, T., K. Watanabe, T. Mie, M. Honma, H. Watanabe, M. Yao, H. Oikawa, and I. Tanaka, Insight into a natural Diels-Alder reaction from the structure of macrophomate synthase. *Nature*, 2003. **422**(6928): 185-9.

154. Townsend, C.A., A "Diels-Alderase" at last. *Chembiochem*, 2011. **12**(15): 2267-9.

155. Kim, H.J., M.W. Ruszczycky, S.H. Choi, Y.N. Liu, and H.W. Liu, Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A. *Nature*, 2011. **473**(7345): 109-12.

156. Podust, L.M. and D.H. Sherman, Diversity of P450 enzymes in the biosynthesis of natural products. *Nat. Prod. Rep.*, 2012. **29**(10): 1251-66.

157. Weber, J.M., J.O. Leung, S.J. Swanson, K.B. Idler, and J.B. McAlpine, An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea*. *Science*, 1991. **252**(5002): 114-7.

158. Arakawa, K., K. Kodama, S. Tatsuno, S. Ide, and H. Kinashi, Analysis of the loading and hydroxylation steps in lankamycin biosynthesis in *Streptomyces rochei*. *Antimicrob. Agents Chemother.*, 2006. **50**(6): 1946-52.

159. Kong, R., X. Liu, C. Su, C. Ma, R. Qiu, and L. Tang, Elucidation of the biosynthetic gene cluster and the post-PKS modification mechanism for fostriecin in *Streptomyces pulveraceus*. *Chem. Biol.*, 2013. **20**(1): 45-54.

160. Niehaus, E.M., K. Kleigrewe, P. Wiemann, L. Studt, C.M. Sieber, L.R. Connolly, M. Freitag, U. Guldener, B. Tudzynski, and H.U. Humpf, Genetic manipulation of the *Fusarium fujikuroi* fusarin gene cluster yields insight into the complex regulation and fusarin biosynthetic pathway. *Chem. Biol.*, 2013. **20**(8): 1055-66.

161. Cane, D.E. and C.C. Yang, Nargenicin biosynthesis: late stage oxidations and absolute configuration. *J. Antibiot.*, 1985. **38**(3): 423-6.

162. Guengerich, F.P. and E.M. Isin, Unusual metabolic reactions and pathways. , in The handbook of metabolic pathways of xenobiotics. Chichester, UK: Wiley; 2014. p. 147-197.

163. Guengerich, F.P. and T.L. Macdonald, Chemical mechanisms of catalysis by cytochromes P-450: a unified view. *Acc. Chem. Res.*, 1984. **17**(1): 9-16.

164. Barry, S.M. and G.L. Challis, Tailoring reactions catalyzed by heme-dependent enzymes: spectroscopic characterization of the L-tryptophan-nitrating cytochrome P450 TxtE. *Methods Enzymol*, 2012. **516**: 171-94.

165. Stassi, D., S. Donadio, M.J. Staver, and L. Katz, Identification of a *Saccharopolyspora erythraea* gene required for the final hydroxylation step in erythromycin biosynthesis. *J. Bacteriol.*, 1993. **175**(1): 182-9.

166. Ruettinger, R.T., L.P. Wen, and A.J. Fulco, Coding nucleotide, 5' regulatory, and deduced amino acid sequences of P-450BM-3, a single peptide cytochrome P-450:NADPH-P-450 reductase from *Bacillus megaterium*. *J. Biol. Chem.*, 1989. **264**(19): 10987-95.

167. Hoffmann, I., F. Jerneren, and E.H. Oliw, Expression of fusion proteins of *Aspergillus terreus* reveals a novel allene oxide synthase. *J. Biol. Chem.*, 2013. **288**(16): 11459-69.

168. Rajgarhia, V.B., N.D. Priestley, and W.R. Strohl, The product of *dpsC* confers starter unit fidelity upon the daunorubicin polyketide synthase of *Streptomyces* sp. strain C5. *Metab. Eng.*, 2001. **3**(1): 49-63.

169. Paitan, Y., E. Orr, E.Z. Ron, and E. Rosenberg, An unusual β-ketoacyl:acyl carrier protein synthase and acyltransferase motifs in TaK, a putative protein required for biosynthesis of the antibiotic TA in *Myxococcus xanthus*. *FEMS Microbiol. Lett.*, 2001. **203**(2): 191-197.

170. Thibodeaux, C.J., C.E. Melancon, 3rd, and H.W. Liu, Natural-product sugar biosynthesis and enzymatic glycodiversification. *Angew. Chem. Int. Ed. Engl.*, 2008. **47**(51): 9814-59.

171. Gao, Q., C. Zhang, S. Blanchard, and J.S. Thorson, Deciphering indolocarbazole and enediyne aminodideoxypentose biosynthesis through comparative genomics: insights from the AT2433 biosynthetic locus. *Chem. Biol.*, 2006. **13**(7): 733-743.

172. Zhang, C., C. Albermann, X. Fu, N.R. Peters, J.D. Chisholm, G. Zhang, E.J. Gilbert, P.G. Wang, D.L. Van Vranken, and J.S. Thorson, RebG- and RebM-catalyzed indolocarbazole diversification. *ChemBioChem*, 2006. **7**(5): 795-804.

173. Zhao, L., N.J. Beyer, S.A. Borisova, and H.W. Liu, Beta-glucosylation as a part of self-resistance mechanism in methymycin/pikromycin producing strain *Streptomyces venezuelae*. *Biochemistry*, 2003. **42**(50): 14794-804.

174. Garrido, L.M., F. Lombo, I. Baig, E.A.M. Nur, R.L. Furlan, C.C. Borda, A. Brana, C. Mendez, J.A. Salas, J. Rohr, and G. Padilla, Insights in the glycosylation steps during biosynthesis of the antitumor anthracycline cosmomycin: characterization of two glycosyltransferase genes. *Appl. Microbiol. Biotechnol.*, 2006. **73**(1): 122-31.

175. Liu, G., K.F. Chater, G. Chandra, G. Niu, and H. Tan, Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol. Mol. Biol. Rev.*, 2013. **77**(1): 112-143.

176. Chater, K. and G. Chandra, The use of the rare UUA codon to define "expression space" for genes involved in secondary metabolism, development and environmental adaptation in *Streptomyces*. *J Microbiol.*, 2008. **46**(1): 1-11.

177. Tao, M., L. Wang, E. Wendt-Pienkowski, N.P. George, U. Galm, G. Zhang, J.M. Coughlin, and B. Shen, The tallysomycin biosynthetic gene cluster from *Streptoalloteichus hindustanus* E465-

94 ATCC 31158 unveiling new insights into the biosynthesis of the bleomycin family of antitumor antibiotics. *Mol. Biosyst.*, 2007. **3**(1): 60-74.

178. Pao, S.S., I.T. Paulsen, and M.H. Saier, Jr., Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.*, 1998. **62**(1): 1-34.

179. Girard, G., J. Willemse, H. Zhu, D. Claessen, K. Bukarasam, M. Goodfellow, and G.P. van Wezel, Analysis of novel kitasatosporae reveals significant evolutionary changes in conserved developmental genes between *Kitasatospora* and *Streptomyces*. *Antonie Van Leeuwenhoek*, 2014. **106**(2): 365-80.

180. Hwang, J.Y., S.H. Kim, H.R. Oh, Y.-J. Cho, J. Chun, Y.R. Chung, and D.H. Nam, Draft genome sequence of *Kitasatospora cheerisanensis* KCTC 2395, which produces plecomacrolide against phytopathogenic fungi. *Genome Announc.*, 2014. **2**(3): e00604-14

181. Cimermancic, P., M.H. Medema, J. Claesen, K. Kurita, L.C. Wieland Brown, K. Mavrommatis, A. Pati, P.A. Godfrey, M. Koehrsen, J. Clardy, B.W. Birren, E. Takano, A. Sali, R.G. Linington, and M.A. Fischbach, Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell*, 2014. **158**(2): 412-21.

182. Djambazian, H., J. Wasserscheid, N. Juretic, G. Geneau, P. Willett , P. Bérubé, G. Leveque, J. Tremblay, L. Létourneau, A. Staffa, R. Sladek, A. Montpetit, and K. Dewar, Achieving complete and accurate assemblies with PacBio RSII and HGAP2. 2014, McGill University and Génome Québec Innovation Centre. 4.

183. Chin, C.-S., D.H. Alexander, P. Marks, A.A. Klammer, J. Drake, C. Heiner, A. Clum, A. Copeland, J. Huddleston, E.E. Eichler, S.W. Turner, and J. Korlach, Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods*, 2013. **10**(6): 563-563.

184. Overbeek, R., R. Olson, G.D. Pusch, G.J. Olsen, J.J. Davis, T. Disz, R.A. Edwards, S. Gerdes, B. Parrello, M. Shukla, V. Vonstein, A.R. Wattam, F. Xia, and R. Stevens, The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*, 2014. **42**(Database issue): D206-14.

185. Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond, Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 2012. **28**(12): 1647-1649.

186. Darling, A.E., B. Mau, and N.T. Perna, progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*, 2010. **5**(6): e11147.

187. Arens, J.C., B. Haltli, and R.G. Kerr, Draft genome sequence of *Kitasatospora griseola* strain MF730-N6, a bafilomycin, terpentecin, and satosporin producer. *Genome Announc.*, 2015. **3**(2): e00208-15.

188. Choulet, F., B. Aigle, A. Gallois, S. Mangenot, C. Gerbaud, C. Truong, F.X. Francou, C. Fourrier, M. Guerineau, B. Decaris, V. Barbe, J.L. Pernodet, and P. Leblond, Evolution of the terminal regions of the *Streptomyces* linear chromosome. *Mol. Biol. Evol.*, 2006. **23**(12): 2361-9.

189. Huang, C.H., Y.S. Lin, Y.L. Yang, S.W. Huang, and C.W. Chen, The telomeres of *Streptomyces* chromosomes contain conserved palindromic sequences with potential to form complex secondary structures. *Mol. Microbiol.*, 1998. **28**(5): 905-16.

190. Jakimowicz, D., J. Majka, W. Messer, C. Speck, M. Fernandez, M.C. Martin, J. Sanchez, F. Schauwecker, U. Keller, H. Schrempf, and J. Zakrzewska-Czerwinska, Structural elements of the *Streptomyces* oriC region and their interactions with the DnaA protein. *Microbiology*, 1998. **144**(Prt 5): 1281-90.

191. Smulczyk-Krawczyszyn, A., D. Jakimowicz, B. Ruban-Osmialowska, A. Zawilak-Pawlik, J. Majka, K. Chater, and J. Zakrzewska-Czerwinska, Cluster of DnaA boxes involved in regulation of *Streptomyces* chromosome replication: from *in silico* to *in vivo* studies. *J. Bacteriol.*, 2006. **188**(17): 6184-94.

192. Chater, K.F. and G. Chandra, The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiol. Rev.*, 2006. **30**(5): 651-672.

193. Eccleston, M., R.A. Ali, R. Seyler, J. Westpheling, and J. Nodwell, Structural and genetic analysis of the BldB protein of *Streptomyces coelicolor*. *J. Bacteriol.*, 2002. **184**(15): 4270-6.

194. Krawczyk, B., J.M. Krawczyk, and R.D. Sussmuth, Class III lantibiotics-an emergining family of thioether-containing peptides, in Drug discovery from natural products, O. Genilloud and F. Vicente, Editors. Dorchester, UK: Royal Soceity of Chemistry; 2012. p. 42-57.

195. Noens, E.E., V. Mersinias, B.A. Traag, C.P. Smith, H.K. Koerten, and G.P. van Wezel, SsgA-like proteins determine the fate of peptidoglycan during sporulation of *Streptomyces coelicolor*. *Mol. Microbiol.*, 2005. **58**(4): 929-44.

196. Aínsa, J., N. Bird, N.J. Ryding, K. Findlay, and K. Chater, The complex *whiJ* locus mediates environmentally sensitive repression of development of *Streptomyces coelicolor* A3(2). *Antonie Van Leeuwenhoek*, 2010. **98**(2): 225-236.

197. van Wezel, G.P., J. van der Meulen, S. Kawamoto, R.G. Luiten, H.K. Koerten, and B. Kraal, *ssgA* is essential for sporulation of *Streptomyces coelicolor* A3(2) and affects hyphal development by stimulating septum formation. *J. Bacteriol.*, 2000. **182**(20): 5653-62.

198. Chater, K.F., Differentiation in *Streptomyces*: the properties and programming of diverse cell-types, in *Streptomyces*: molecular biology and biotechnology, P. Dyson, Editor. Norfolk, UK: Caister Academic Press; 2011. p. 43-86.

199. Welander, P.V., D.M. Doughty, C.H. Wu, S. Mehay, R.E. Summons, and D.K. Newman, Identification and characterization of *Rhodopseudomonas palustris* TIE-1 hopanoid biosynthesis mutants. *Geobiol.*, 2012. **10**(2): 163-77.

200. Jiaoyang, J., H. Xiaofei, and D.E. Cane, Biosynthesis of the earthy odorant geosmin by a bifunctional *Streptomyces coelicolor* enzyme. *Nat. Chem. Biol.*, 2007. **3**(11): 711-715.

201. Cane, D.E. and H. Ikeda, Exploration and mining of the bacterial terpenome. *Acc. Chem. Res.*, 2012. **45**(3): 463-72.

202. Hider, R.C. and X. Kong, Chemistry and biology of siderophores. *Nat. Prod. Rep.*, 2010. **27**(5): 637-657.

203. Challis, G.L., A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. *ChemBioChem*, 2005. **6**(4): 601-611.

204. Dobson, A., P.D. Cotter, R.P. Ross, and C. Hill, Bacteriocin production: a probiotic trait? *Appl. Environ. Microbiol.*, 2012. **78**(1): 1-6.

205. Valdes-Stauber, N. and S. Scherer, Nucleotide sequence and taxonomical distribution of the bacteriocin gene *lin* cloned from *Brevibacterium linens* M18. *Appl. Environ. Microbiol.*, 1996. **62**(4): 1283-6.

206. Yamato, M., H. Iinuma, H. Naganawa, Y. Yamagishi, M. Hamada, T. Masuda, H. Umezawa, V. Abe, and M. Hori, Isolation and properties of valanimycin, a new azoxy antibiotic. *J. Antibiot.*, 1986. **39**(2): 184-91.

207. Tiwari, K. and R.K. Gupta, Rare actinomycetes: a potential storehouse for novel antibiotics. *Crit. Rev. Biotechnol.*, 2012. **32**(2): 108-132.

208. Pluskal, T., S. Castillo, A. Villar-Briones, and M. Oresic, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, 2010. **11**: 395.

209. Uramoto, M., Y. Itoh, R. Sekiguchi, K. Shin-Ya, H. Kusakabe, and K. Isono, A new antifungal antibiotic cystargin. Fermentation, isolation and characterization. *J. Antibiot.*, 1988. **41**(12): 1763-1768.

210. Gill, K.A., F. Berrue, J.C. Arens, G. Carr, and R.G. Kerr, Cystargolides, 20S proteasome inhibitors isolated from *Kitasatospora cystarginea*. *J. Nat. Prod.*, 2015. **78**(4): 822-826.

211. Gill, K.A., F. Berrue, J.C. Arens, and R.G. Kerr, Isolation and structure elucidation of cystargamide, a lipopeptide from *Kitasatospora cystarginea*. *J. Nat. Prod.*, 2014. **77**(6): 1372-6.

212. Umezawa, H., T. Aoyagi, H. Morishima, M. Matsuzaki, and M. Hamada, Pepstatin, a new pepsin inhibitor produced by actinomycetes. *J. Antibiot.*, 1970. **23**(5): 259-62.

213. Balerna, M., W. Keller-Schierlein, C. Martius, H. Wolf, and H. Zähner, Stoffwechselprodukte von mikroorganismen. *Arch. Mikrobiol.*, 1969. **65**(4): 303-317.

214. Yoon, T.M., J.W. Kim, J.G. Kim, W.G. Kim, and J.W. Suh, Talosins A and B: new isoflavonol glycosides with potent antifungal activity from *Kitasatospora kifunensis* MJM341 I. Taxonomy, fermentation, isolation, and biological activities. *J. Antibiot.*, 2006. **59**(10): 633-639.

215. Rajamani, S., W.D. Bauer, J.B. Robinson, J.M. Farrow, 3rd, E.C. Pesci, M. Teplitski, M. Gao, R.T. Sayre, and D.A. Phillips, The vitamin riboflavin and its derivative lumichrome activate the LasR bacterial quorum-sensing receptor. *Mol. Plant. Microbe Interact.*, 2008. **21**(9): 1184-92.

216. Phillips, D.A., C.M. Joseph, G.P. Yang, E. Martinez-Romero, J.R. Sanborn, and H. Volpin, Identification of lumichrome as a sinorhizobium enhancer of alfalfa root respiration and shoot growth. *Proc. Natl. Acad. Sci. U. S. A.*, 1999. **96**(22): 12275-80.

217. Kurtboeke, D.I., Biodiscovery from rare actinomycetes: an eco-taxonomical perspective. *Appl. Microbiol. Biotechnol.*, 2012. **93**(5): 1843-1852.

APPENDIX A - SUPPORTING INFORMATION REGARDING THE STRUCTURE ELUCIDATION OF SATOSPORIN A-C (CHAPTER 3)

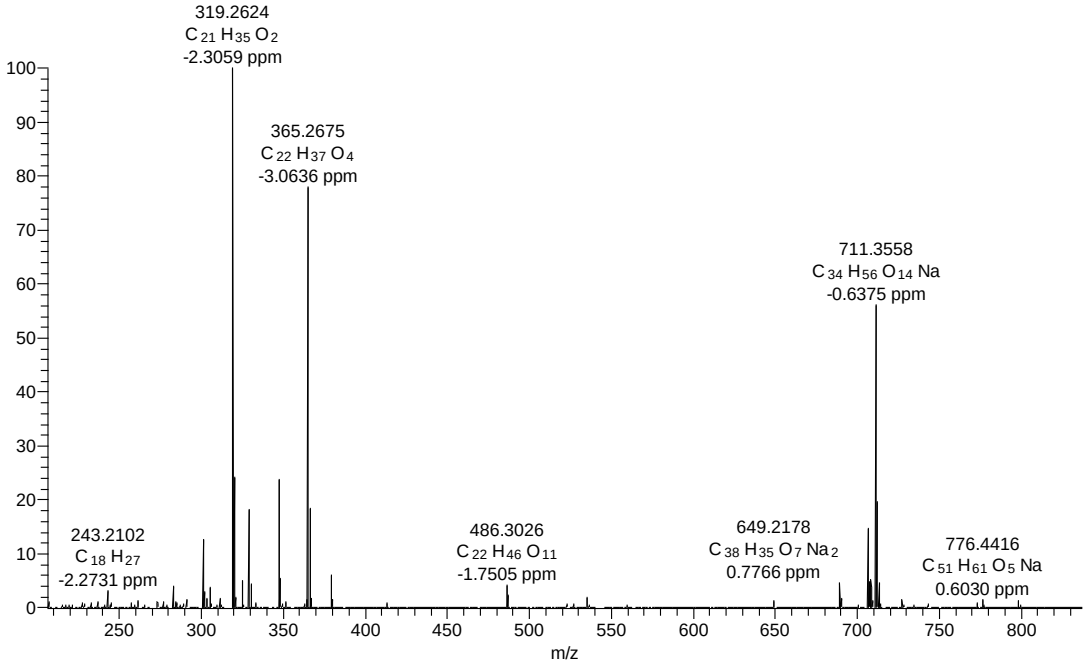


Figure A.1. HRMS of satosporin A.

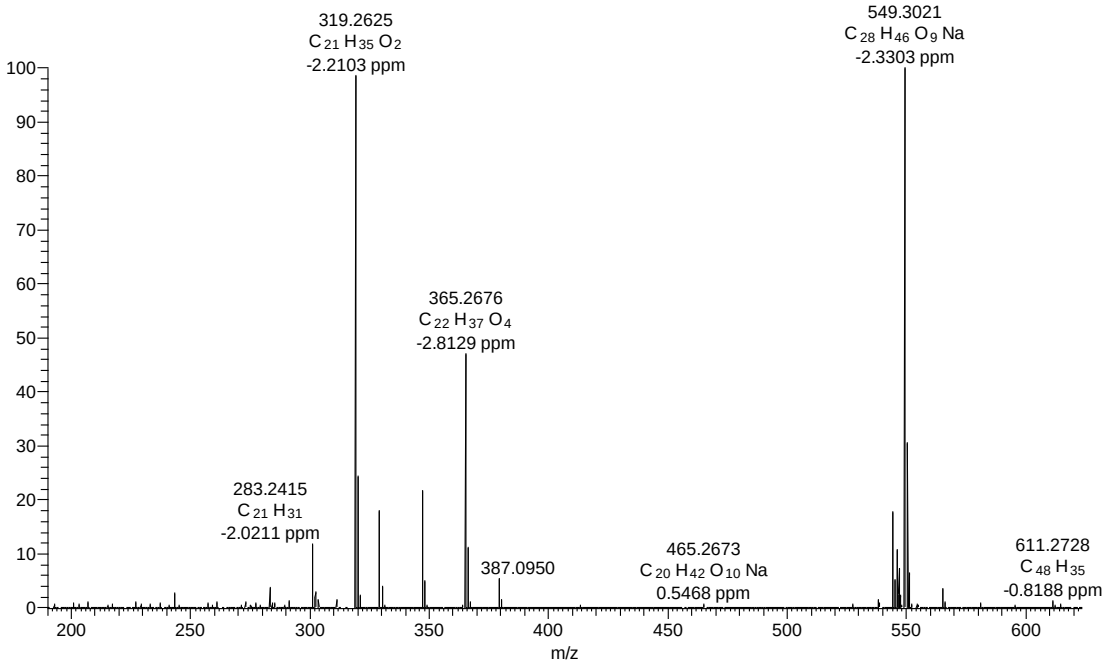


Figure A.2. HRMS of satosporin B.

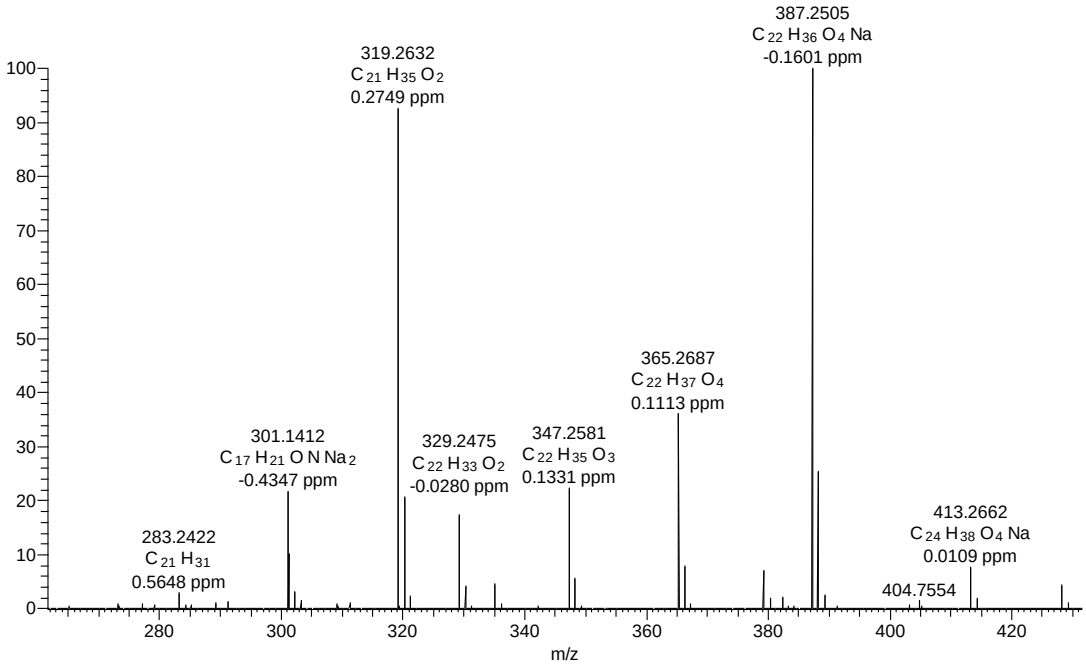


Figure A.3. HRMS of satosporin C.

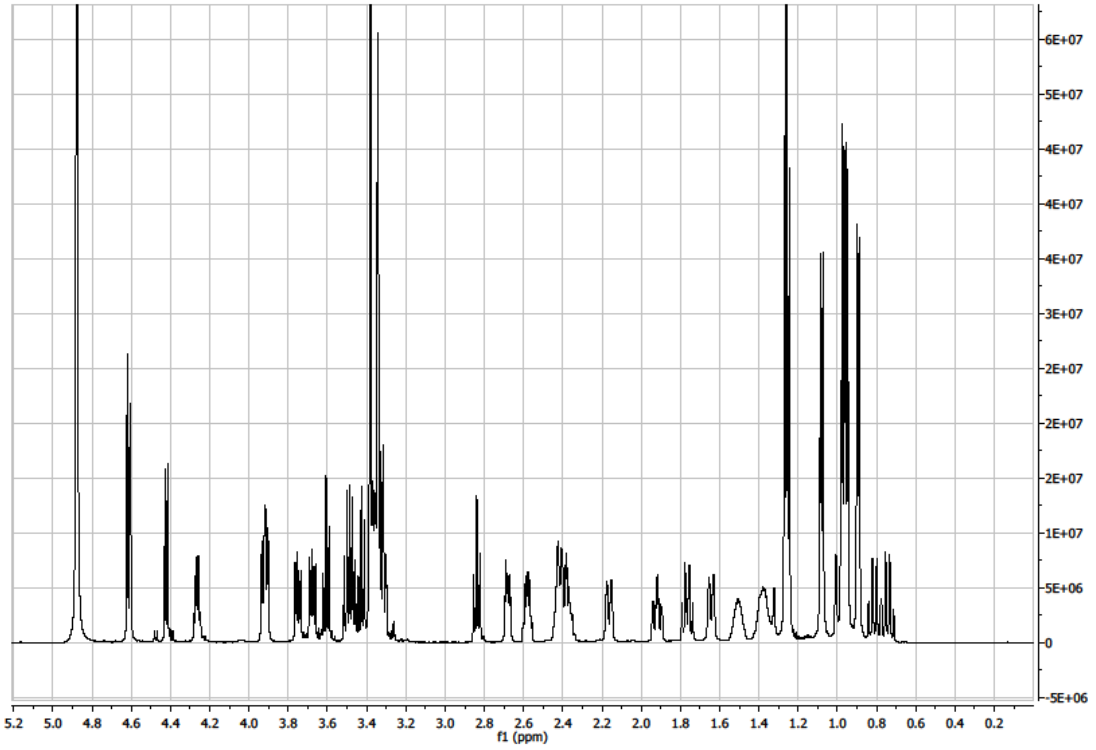


Figure A.4. ^{1}H NMR spectrum (600 MHz, CD_3OD) of satosporin A.

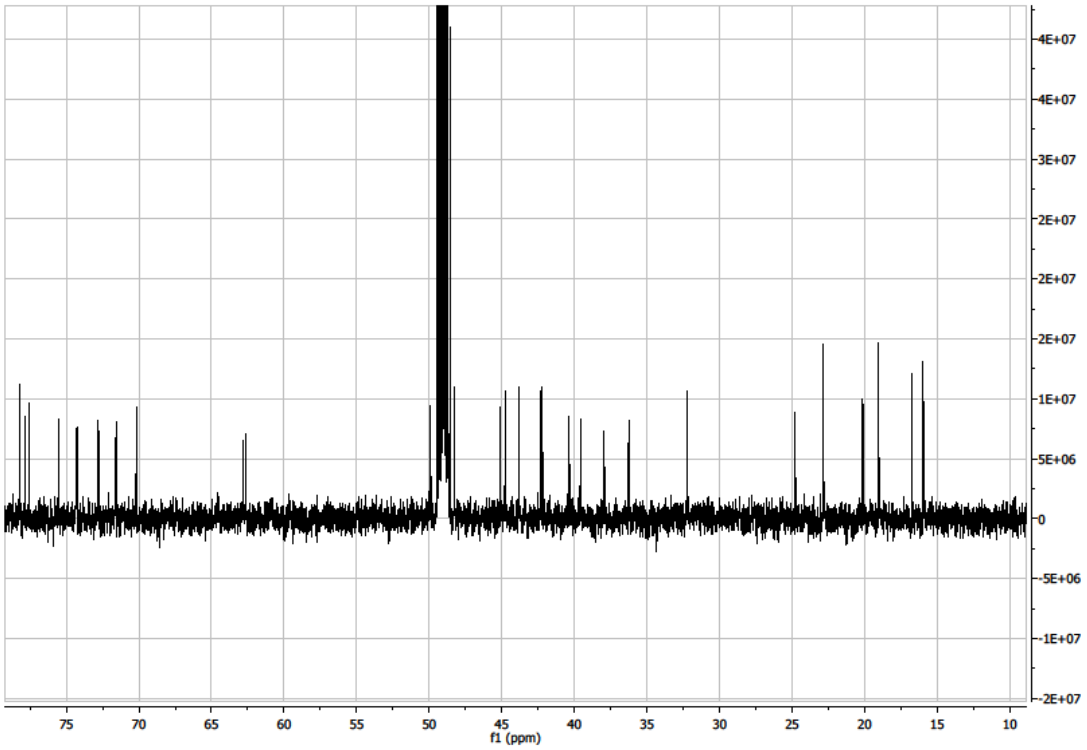


Figure A.5. ^{13}C NMR spectrum (150 MHz, CD_3OD) of satosporin A.

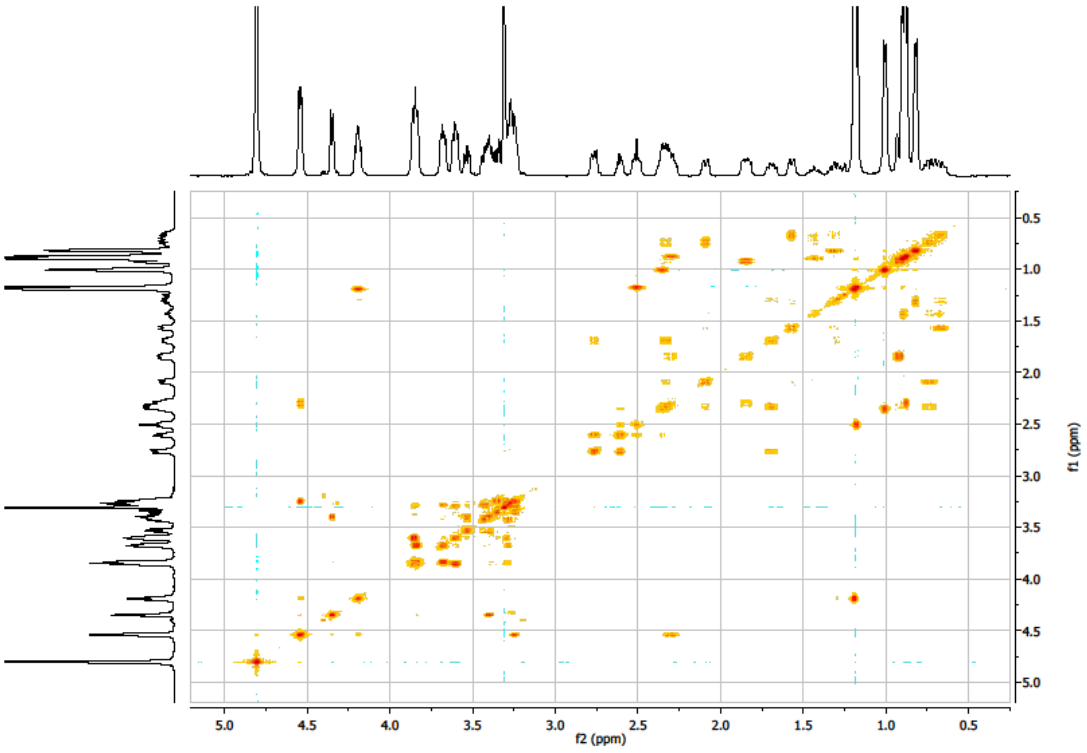


Figure A.6. COSY spectrum of satosporin A.

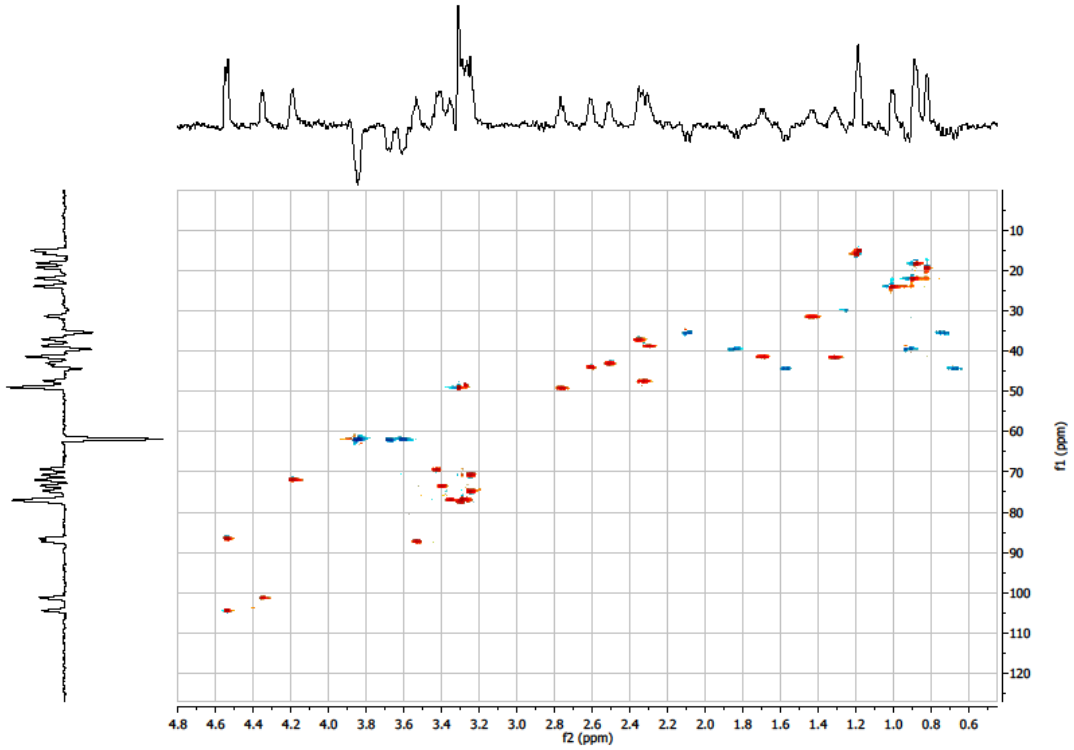


Figure A.7. HSQC spectrum of satosporin A.

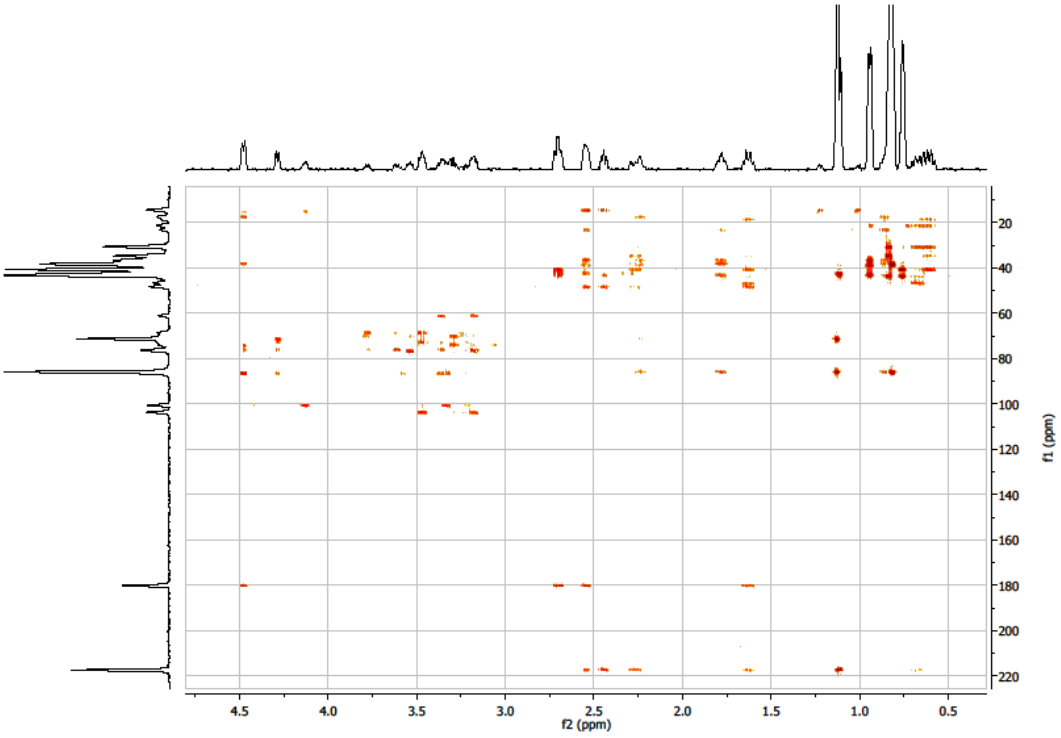


Figure A.8. HMBC spectrum (600 MHz, CD_3OD) of satosporin A.

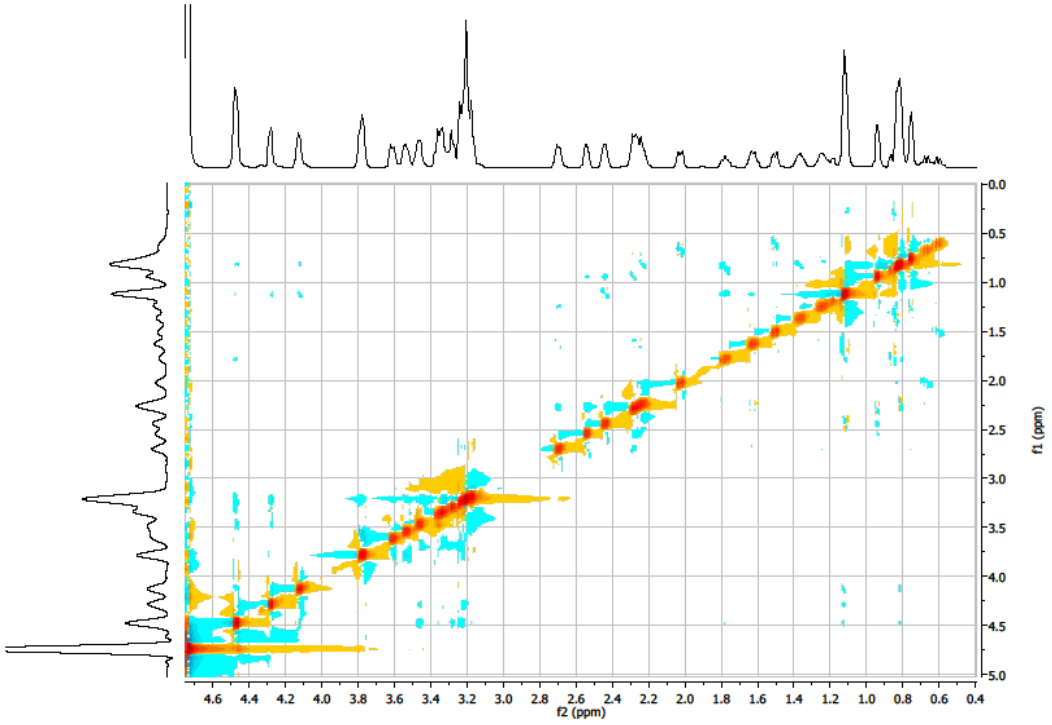


Figure A.9. ROESY spectrum (600 MHz, CD_3OD) of satosporin A.

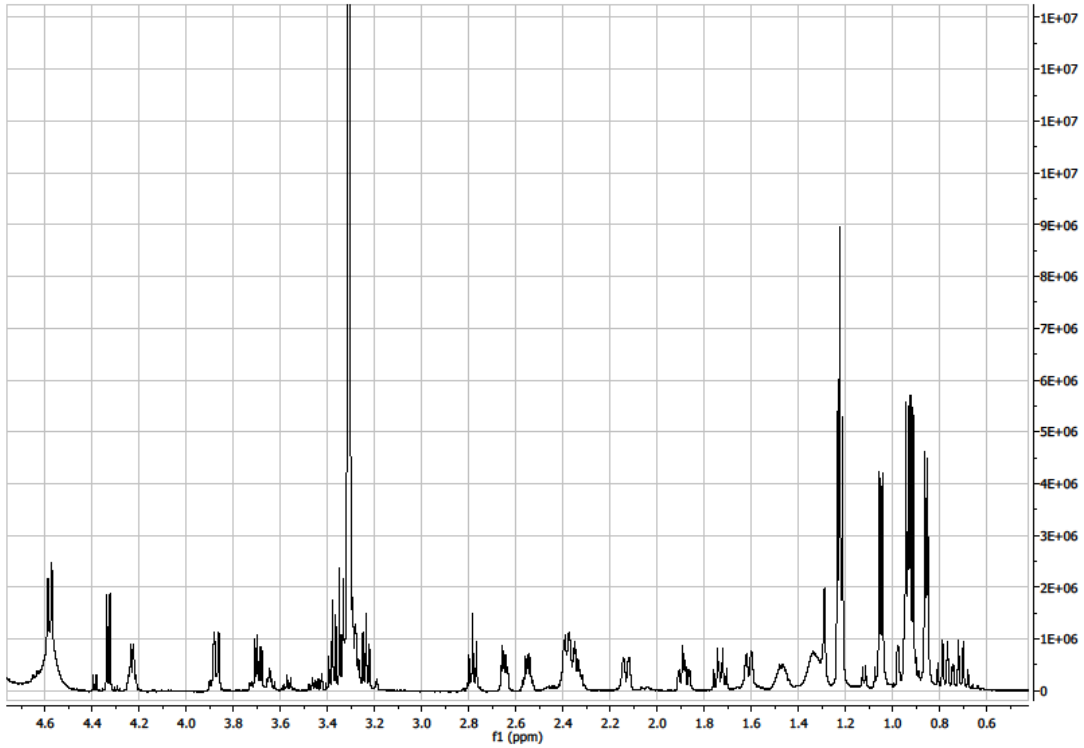


Figure A.10. ^1H NMR spectrum (600 MHz, CD_3OD) of satosporin B.

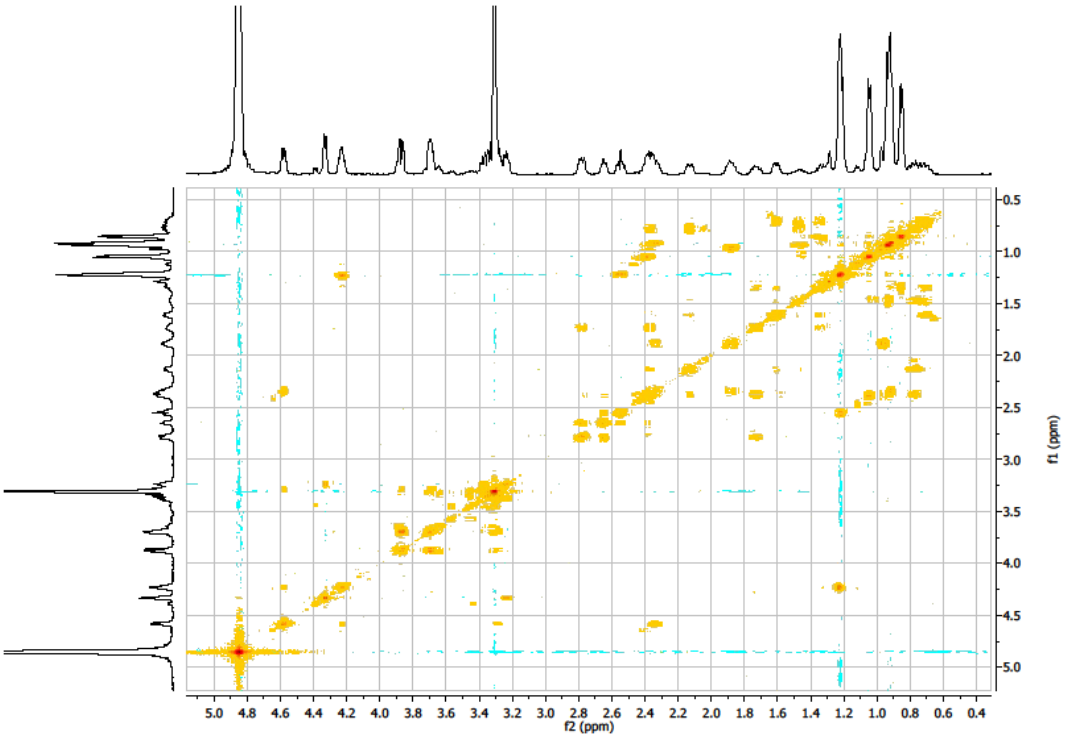


Figure A.11. COSY spectrum of satosporin B.

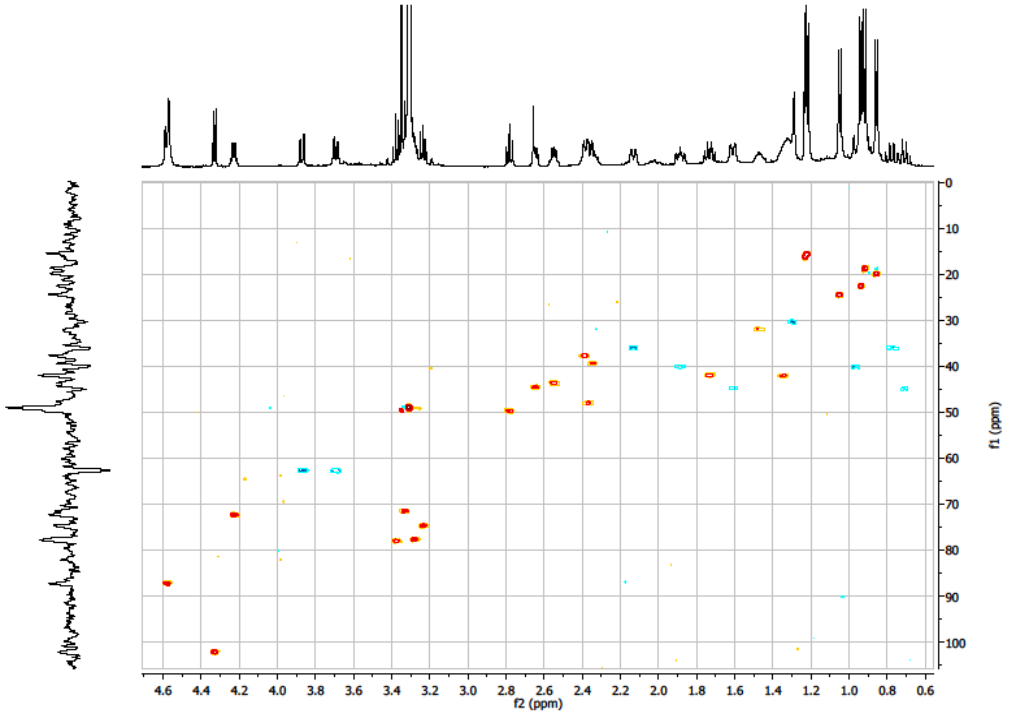


Figure A.12. HSQC spectrum (600 MHz, CD_3OD) of satosporin B.

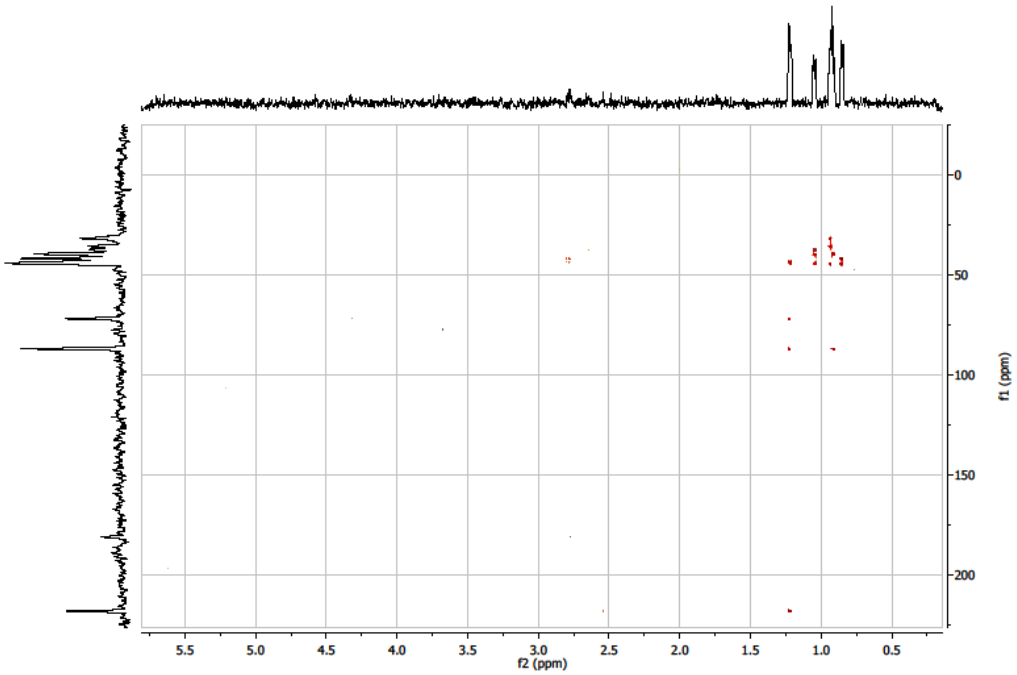


Figure A.13. HMBC spectrum (600 MHz, CD_3OD) of satosporin B.

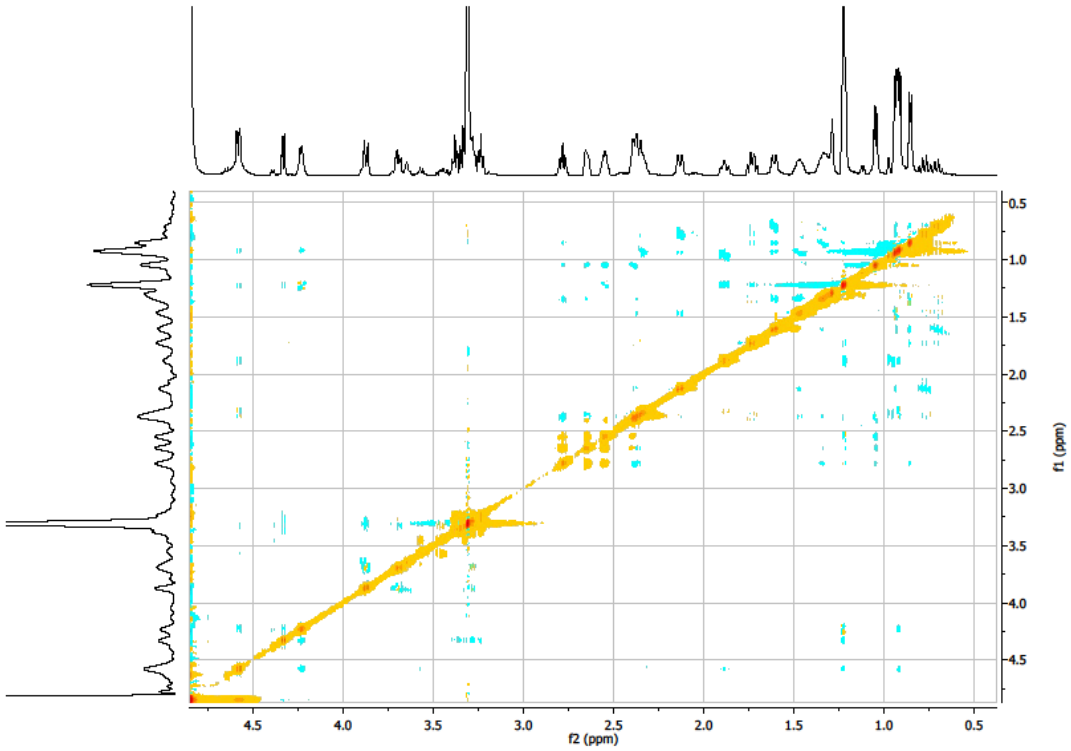


Figure A.14. ROESY spectrum of satosporin B.

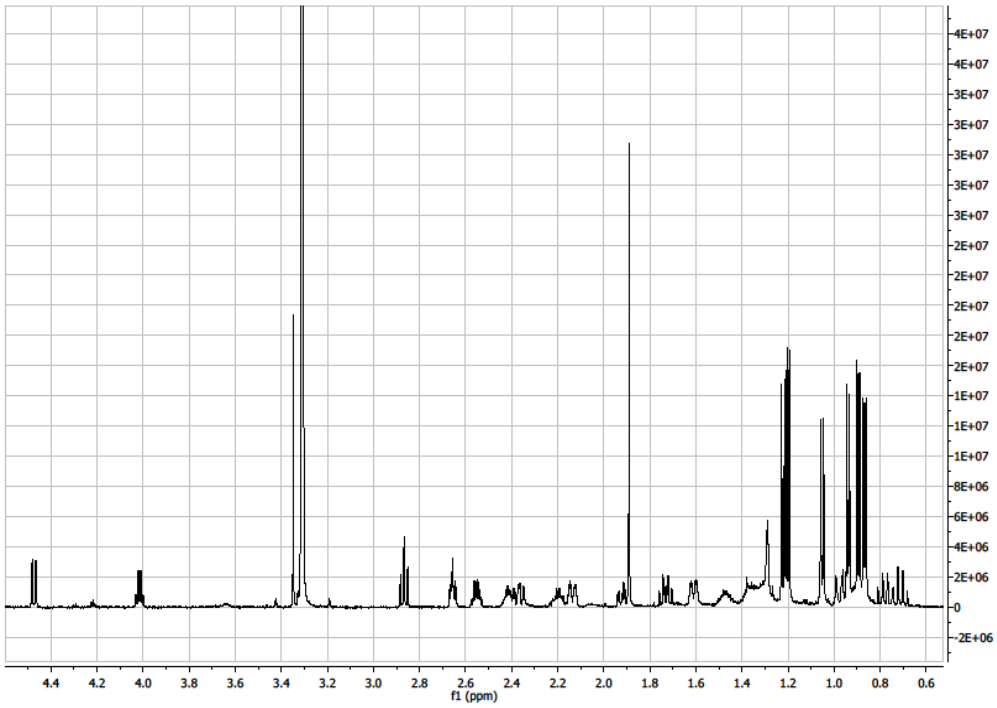


Figure A.15. ^{1}H NMR spectrum (600 MHz, CD_3OD) of satosporin C.

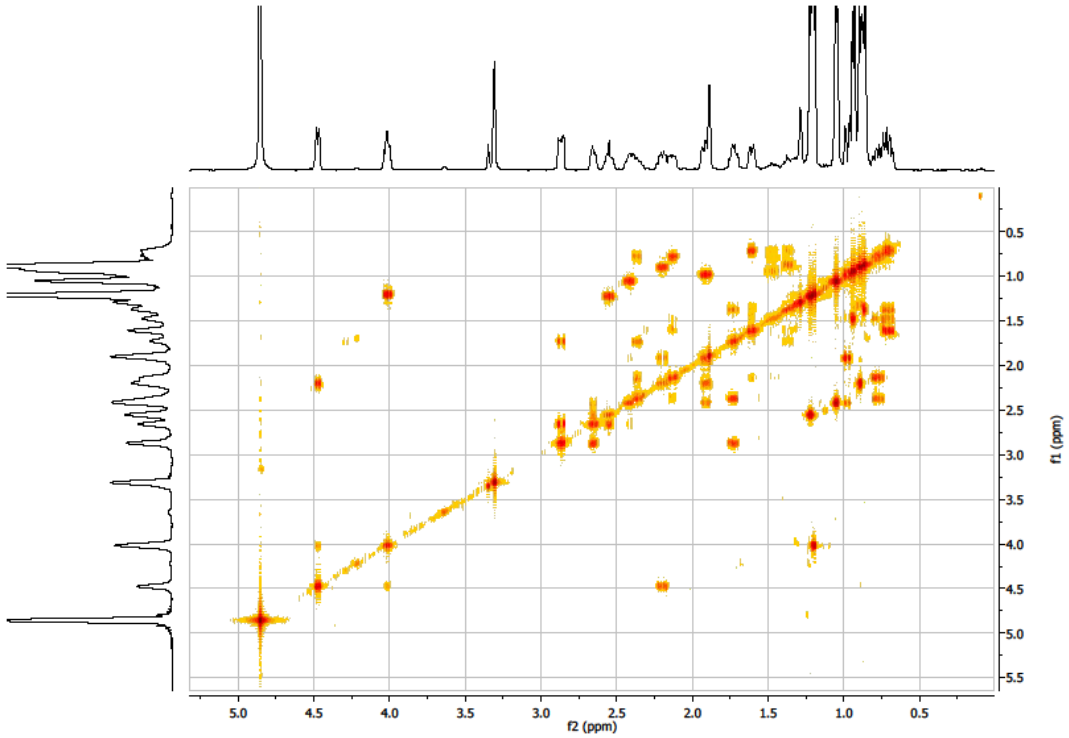


Figure A.16. COSY spectrum of satosporin C.

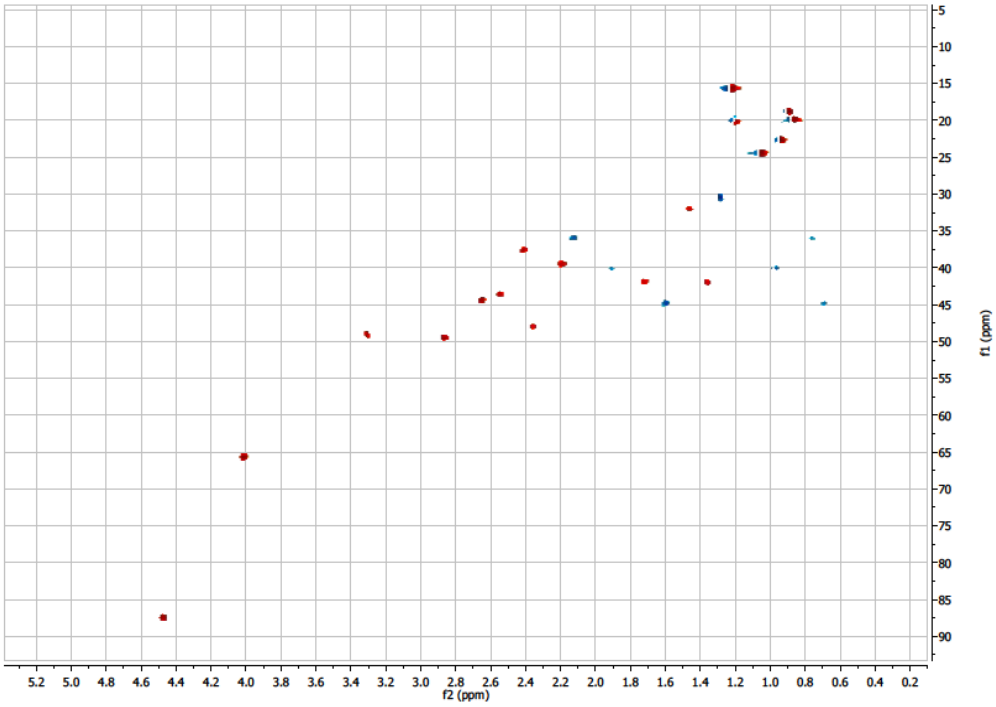


Figure A.17. HSQC spectrum of satosporin C.

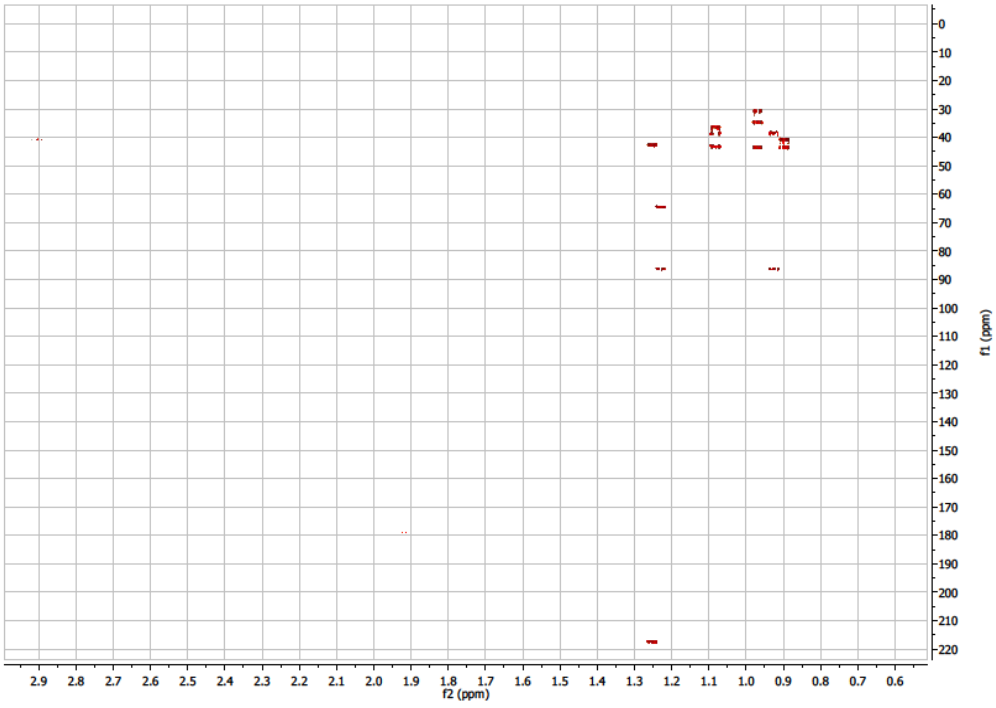


Figure A.18. HMBC spectrum of satosporin C.

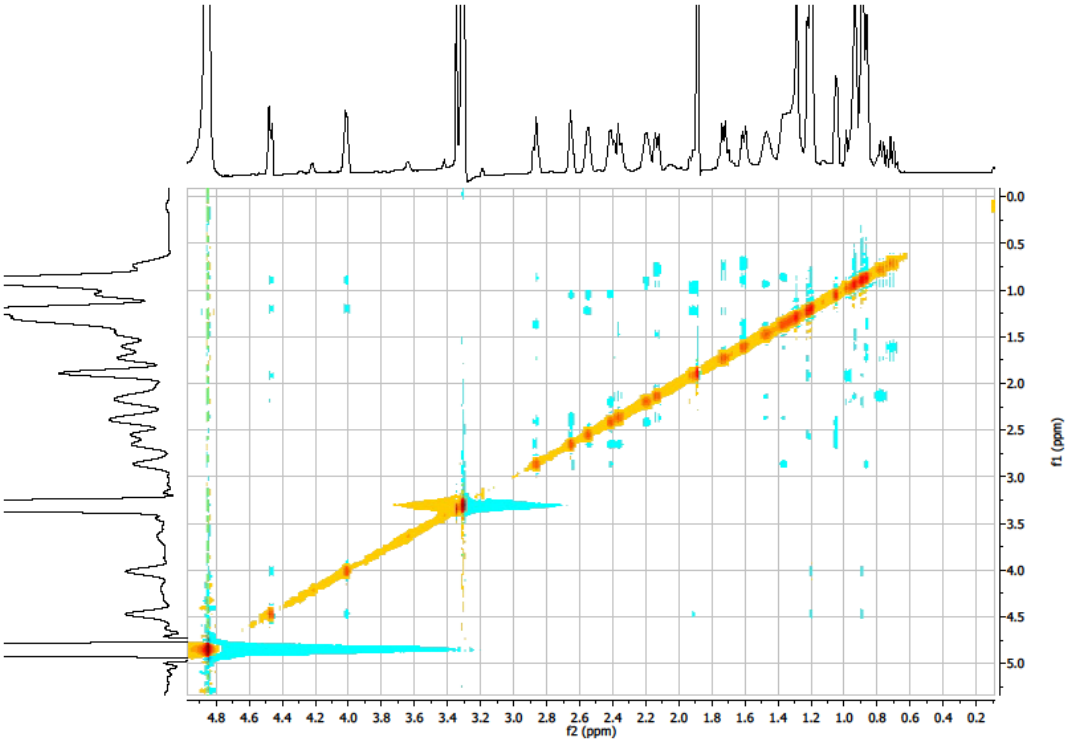


Figure A.19. NOESY spectrum of satosporin C.

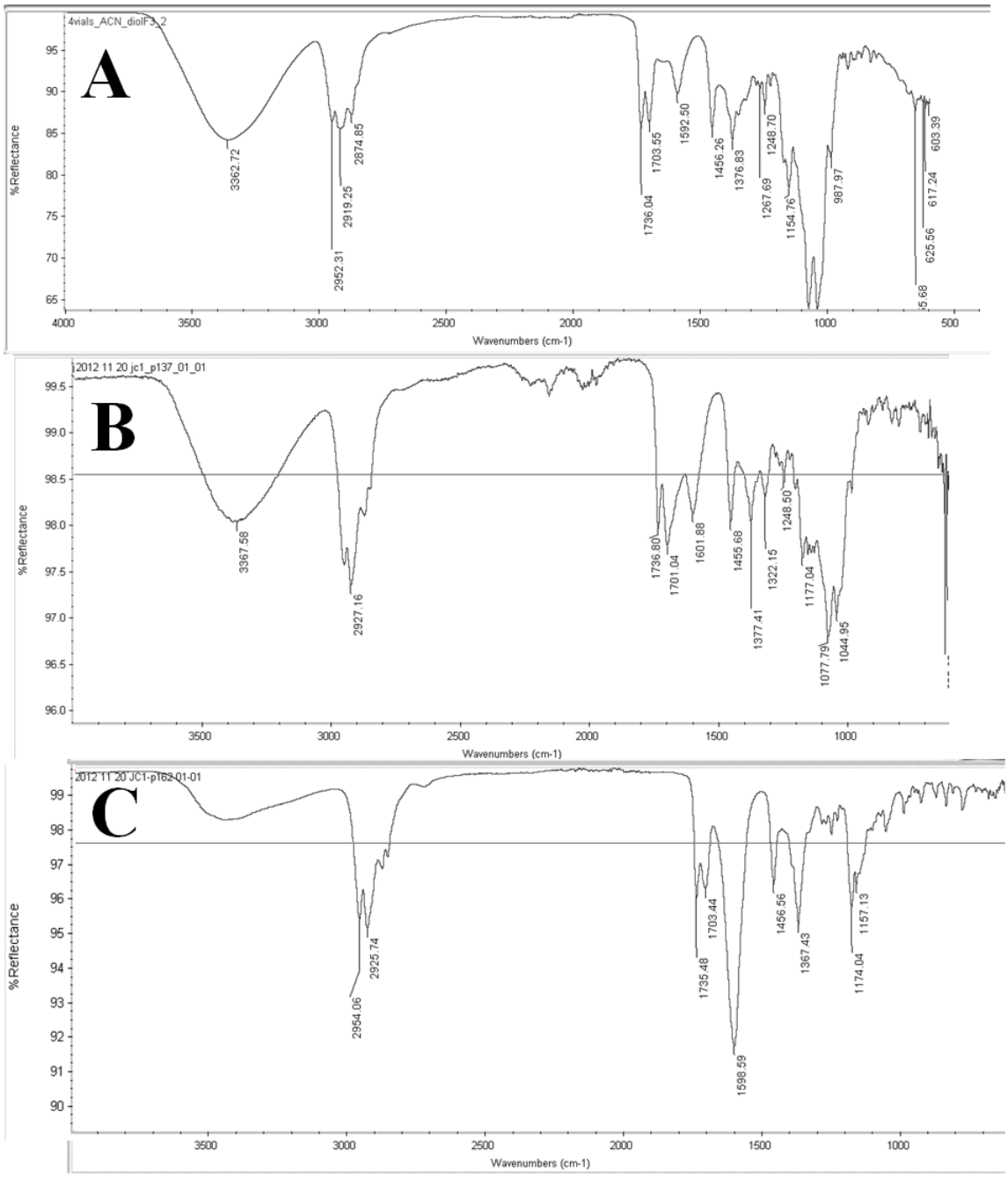


Figure A.20. Infrared spectra (IR) recorded at 25°C in methanol for satosporin A **(A)**, satosporin B **(B)** and satosporin C **(C)**.